Will fusion power run out of fuel before it even gets started? p. 1372

Building amines from nitriles pp. 1382 & 1433

Most turtles and tortoises age slowly pp. 1384, 1459, & 1466

Science

$15
24 JUNE 2022
SPECIAL ISSUE
science.org
AAAS

A CHOICE OF FUTURES

Humanity can still avert climate catastrophe p. 1392

CONTENTS

SPECIAL SECTION

Our climate future

INTRODUCTION

1392 Time to act

NEWS

1394 Can biofuels really fly? *R. F. Service*
PODCAST

POLICY FORUMS

1398 Strengthen climate adaptation research globally *A. K. Magnan* et al.

1401 How trade policy can support the climate agenda *M. Jakob* et al.

REVIEWS

1404 Current global efforts are insufficient to limit warming to 1.5°C *H. D. Matthews and S. Wynes*

1410 Harnessing the potential of nature-based solutions for mitigating and adapting to climate change *N. Seddon*
PODCAST

1416 Climate change and the urgency to transform food systems *M. Zurek* et al.

1421 Getting ahead of climate change for ecological adaptation and resilience *J. W. Moore and D. E. Schindler*

ON THE COVER

Humanity's actions have committed us to a warming climate and limited our options for mitigation. Although there is no turning back, some paths are still open to avoid catastrophic climate change and reduce its impacts. We must act now to drastically reduce greenhouse gas emissions and change our approaches to growing food, consuming products, and managing ecosystems to avoid a dire future. See page 1392. *Illustration: Myriam Wares*

SEE ALSO EDITORIAL p. 1361
NEWS STORY p. 1372 PODCAST

NEWS

IN BRIEF

1362 News at a glance

IN DEPTH

1364 Why monkeypox is mostly hitting men who have sex with men
The once slow spreading virus may have found a new niche in tightly connected sexual networks *By K. Kupferschmidt*

1365 Stimulating the brain may help people who stutter
Trials suggest small electric currents aid fluency, but how long the benefits last is unclear
By I. Amigo

1367 Arati Prabhakar to lead Biden's science office
Applied physicist would bring wealth of policy experience as successor to Eric Lander
By J. Mervis

1368 HIV prevention tool is too slow to reach Africa, activists say
A long-lasting injectable drug was tested in seven African countries. But it's unclear when they'll have access to it *By A. Green*

1369 U.S. labs face severe postdoc shortage
Early-career researchers increasingly avoid low-paying, insecure postdoc positions, a *Science* investigation finds *By K. Langin*

1370 Disgraced stem cell surgeon convicted of criminal harm
Swedish court gives Paolo Macchiarini suspended sentence for death of windpipe transplant patient *By G. Vogel*

FEATURES

1372 Out of gas
A shortage of tritium fuel may leave fusion energy with an empty tank *By D. Clery*

1376 Breakdowns could plague fusion power plants
By D. Clery
OUR CLIMATE FUTURE SECTION p. 1392

INSIGHTS

PERSPECTIVES

1378 Phytoplankton response to a warming ocean
The nutritional value of marine algae will decrease in a warmer world
By J. Sepúlveda and S. I. Cantarero
REPORT p. 1487

1379 A bacterium that is not a microbe
A new discovery challenges the prevailing view of the boundaries of bacterial cell size
By P. A. Levin
RESEARCH ARTICLE p. 1453

1381 The many faces of anthropogenic subsidence
Shallow and deep human subsurface activities contribute to the total subsidence
By T. Candela and K. Koster

ILLUSTRATION: MYRIAM WARES

1382 Nitriles for the production of various amines
Nitriles functionalize amines and ammonia under catalytic hydrogenation conditions
By Y. Monguchi
RESEARCH ARTICLE p. 1433

1384 How ubiquitous is aging in vertebrates?
Two new studies find little evidence of aging in some turtle species *By S. N. Austad and C. E. Finch*
RESEARCH ARTICLE p. 1459; REPORT p. 1466

1385 The delicate balance of river sediments
Global satellite data quantify changes in sediment flux in 414 rivers *By C. Zarfl and F. E. Dunn*
RESEARCH ARTICLE p. 1447

BOOKS *ET AL.*

1387 Alphabetized
The origins of the world's chief writing system come to life in an illuminating historiography
By A. Robinson

1388 Dispatches from the redwood rebellion
Tourism fails to compensate for lost timber jobs in protected forests, leading some to theft *By J. B. Yoder*

LETTERS

1389 A wrong turn on roads in Brazil's national parks
By L. R. Forti et al.

1389 Diplomacy for the world's hottest sea
By N. Al-Mudaffar Fawzi et al.

1390 Minimize food loss and waste to prevent crises *By S. You* et al.

1390 Technical Comment abstracts

RESEARCH

IN BRIEF

1428 From *Science* and other journals

RESEARCH ARTICLES

1431 Immunometabolism
Tumor necrosis factor induces pathogenic mitochondrial ROS in tuberculosis through reverse electron transport
F. J. Roca et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABH2841

1432 Coronavirus
Twin peaks: The Omicron SARS-CoV-2 BA.1 and BA.2 epidemics in England
P. Elliott et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABQ4411

1433 Organic chemistry
Nickel-catalyzed hydrogenative coupling of nitriles and amines for general amine synthesis
V. G. Chandrashekhar et al.
PERSPECTIVE p. 1382

1441 Neurodevelopment
Fate mapping of neural stem cell niches reveals distinct origins of human cortical astrocytes
D. E. Allen et al.

1447 Geomorphology
Rapid changes to global river suspended sediment flux by humans
E. N. Dethier et al.
PERSPECTIVE p. 1385

1453 Microbiology
A centimeter-long bacterium with DNA contained in metabolically active, membrane-bound organelles
J.-M. Volland et al.
PERSPECTIVE p. 1379

1459 Aging
Diverse aging rates in ectothermic tetrapods provide insights for the evolution of aging and longevity *B. A. Reinke* et al.
PERSPECTIVE p. 1384; REPORT p. 1466

REPORTS

1466 Aging
Slow and negligible senescence among testudines challenges evolutionary theories of senescence *R. da Silva* et al.
PERSPECTIVE p. 1384; RESEARCH ARTICLE p. 1459

1471 DNA damage repair
Small-molecule activation of OGG1 increases oxidative DNA damage repair by gaining a new function
M. Michel et al.

1476 CRISPR
Structural basis for RNA-guided DNA cleavage by IscB-ωRNA and mechanistic comparison with Cas9
G. Schuler et al.

1481 Organic chemistry
Catalyst-controlled site-selective methylene C–H lactonization of dicarboxylic acids
H. S. S. Chan et al.

1487 Marine lipids
Global ocean lipidomes show a universal relationship between temperature and lipid unsaturation *H. C. Holm* et al.
PERSPECTIVE p. 1378

1491 Forest ecology
8000-year doubling of Midwestern forest biomass driven by population- and biome-scale processes *A. M. Raiho* et al.

DEPARTMENTS

1361 Editorial
The matter of a clean energy future
By J. M. Turner
OUR CLIMATE FUTURE SECTION p. 1392

1498 Working Life
Full circle *By P. Religia*

Science Careers 1496

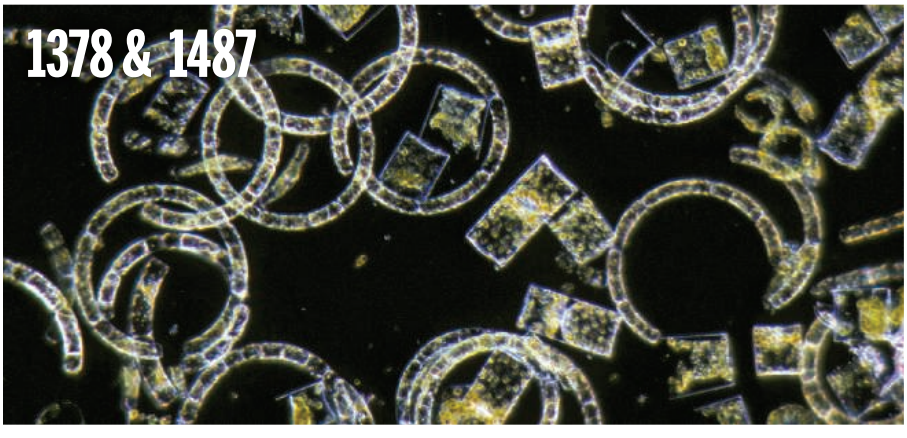


PHOTO: INGO ARNDT/MINDEN PICTURES

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2022 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2212; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

EDITORIAL

The matter of a clean energy future

A clean energy transition will create jobs, promote energy independence, improve public health, and, ultimately, mitigate climate change. But getting to this new future will require more than just phasing out fossil fuels. The production of a wide range of energy-relevant materials—lithium, cobalt, and nickel for batteries; rare earth elements for wind turbines and electric motors; silicon for solar panels; and copper to expand the electric grid—must be scaled up substantially. Mobilizing these materials without reproducing the environmental harms and social inequities of the fossil fuel status quo poses an urgent challenge.

Studies project that producing the materials to enable a clean energy transition will be a massive undertaking. The International Energy Agency forecasts that keeping the world on a path compatible with the goals of the Paris Climate Accord will require expanding production of energy-relevant materials sixfold between 2020 and 2040, to 43 million tons per year. At first glance, that may seem to pale in comparison to the fossil fuel industries, which produced roughly 15 billion tons of coal, oil, and natural gas globally in 2020 alone and added 32 billion tons of carbon dioxide to the atmosphere when burned.

> "...producing the materials to enable a clean energy transition will be a massive undertaking."

But the transition will be even more difficult than it first appears. Nickel, cobalt, and copper and many other energy-relevant materials occur in low-grade ores, which entail far more mining, processing, and waste than fossil fuels. Securing the millions of tons of finished materials needed will require mining hundreds or thousands of times more raw ore. Although this transition will ultimately lower greenhouse gas emissions, especially as more renewable energy powers mining processes, it will require processing metal ores at a scale that rivals the material throughput of today's fossil fuel industries.

The potential harms of such a transition are considerable. Large-scale mining affects ecosystems, threatens water supplies, and is sometimes linked to poor working conditions, corruption, and human rights abuses. But scaling up mining to support a clean energy transition also offers the opportunity to reform materials production in ways that are both socially and environmentally just. Wealthier countries, which have often outsourced mineral extraction abroad, need to help shoulder these burdens and model responsible approaches to development.

To meet the global clean energy challenge, government policies supporting public and private sector investments are needed at every stage of extraction and processing. This means support for exploration, research into new mining and processing technologies, streamlined permitting processes, support for expanding processing capacity, and trade agreements that ensure supplies from international markets. Over the past decade, China has prioritized such public-private investments from mine to factory. Only recently have the United States and European Union begun to adopt similar policies aimed at developing domestic supply chains and diversifying international sources of energy-relevant materials.

These policies must be paired with initiatives to ensure that materials are sourced sustainably and transparently. Over the past decade, third-party certification programs have proliferated in the mining sector. One of the most promising approaches is a standard championed by the Initiative for Responsible Mining Assurance. It offers an independent certification system that can be used to assess mining activities relative to best practices for worker health and safety, human rights, community engagement, corruption, pollution control, and land reclamation through mandatory third-party audits and publicly available scorecards.

But if such standards are going to be effective, policymakers, corporations, investors, environmental groups, and consumers must demand that they be integrated into the policies that support new mines, govern mining operations, and regulate international trade. Such approaches will be essential to ensuring that as extractive industries scale up, information about the impacts of their activities is available and private actors who align their operations with social and sustainability priorities are rewarded.

Ultimately, innovation will reshuffle the burdens of resource extraction in ways that cannot be fully anticipated. Already there are efforts to reduce or, in some cases, eliminate the use of cobalt in lithium-ion batteries. But innovation will not eliminate the material-intensive nature of clean energy technologies. Given the urgent need to mitigate climate change and the long lead time for developing mines and building out supply chains, the time to reckon with the material demands of a clean energy transition is now.

–James Morton Turner

PHOTO: ADAM PULZETTI

James Morton Turner
is a professor of environmental studies at Wellesley College, Wellesley, MA, USA. jturner@wellesley.edu

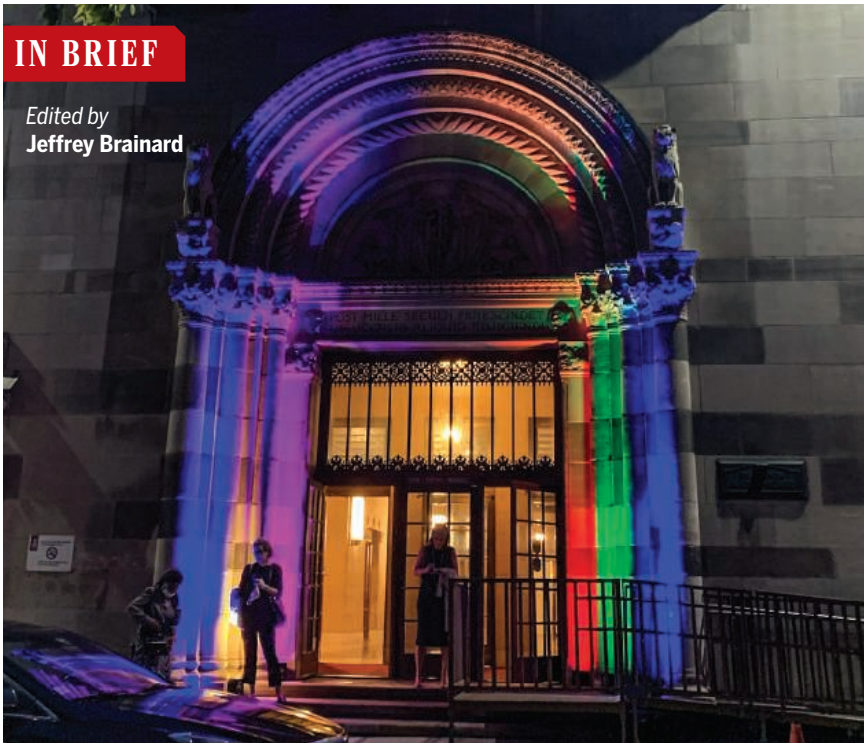
10.1126/science.add5094

NEWS

> “We don’t want to get to a point where we say based on your brain chart, you’re not qualified for a job.”
>
> **Neuroethicist Laura Cabrera,** in *STAT*, on ways that new reference charts describing brain development, published in *Nature*, could fuel discrimination if misused.

IN BRIEF

Edited by
Jeffrey Brainard



The New York Academy of Medicine lit its entrance to celebrate Pride Month, and offered an apology.

PHOTO: NEW YORK ACADEMY OF MEDICINE

DIVERSITY

Groups regret ‘homosexuality’ views

Two scientific societies this month disavowed their past involvement in practices and public statements that deemed “homosexuality” a treatable disorder—a mistaken notion that has harmed LGBTQI+ people. Decades ago, some members and former presidents of the Association for Behavioral and Cognitive Therapies helped create, study, and use “conversion therapies” for sexual and gender minorities. In a statement, the group apologized for its members’ involvement with these practices and accepted responsibility for harm caused. Conversion therapy—whose practitioners attempt to change a person’s sexual orientation or gender identity—has been proven ineffective and is associated with an increased risk of suicide attempts. The association encouraged its members to speak out against these practices in U.S. states that still haven’t banned them. And for this year’s Pride Month, in June, the New York Academy of Medicine (NYAM) formally disowned and apologized for a 1964 report that incorrectly concluded that homosexuality was a “treatable illness.” NYAM’s president called the long delay in issuing the disavowal “shameful.”

Racial gap in NIH funding narrows

BIOMEDICINE | The success rate for Black scientists seeking research grants from the U.S. National Institutes of Health (NIH) has risen sharply the past 2 years, narrowing but not eliminating a gap with white scientists, agency officials said last week. In the 2021 fiscal year, a Black applicant’s odds of receiving at least one new R01, NIH’s standard research grant, was 24.4%, or 2.2 percentage points lower than for a white applicant. That compares with a disparity of about 7 to 9 percentage points annually from 2013 to 2019. A 2011 study identifying the funding gap led NIH to expand training and mentoring for Black researchers; advocates have also urged the agency to fund Black scientists’ applications that just miss the quality score cutoff for funding. Another new effort could help further close the gap: This month, several NIH institutes launched an R01 program for new investigators from diverse backgrounds, limited to research in neuroscience, drug abuse, and mental health. NIH defines “diverse” broadly, to include for example disabled scientists and those from impoverished backgrounds. Agency officials noted that NIH cannot make funding decisions on the basis of race, gender, or ethnicity.

Long Covid grows less likely

PUBLIC HEALTH | People were less than half as likely to develop Long Covid after being infected with the Omicron variant of SARS-CoV-2 than the earlier Delta variant, according to a large epidemiological study. Researchers at King’s College London analyzed data from the COVID Symptom Study, a longitudinal study in which initially healthy participants voluntarily log symptoms and COVID-19 test results on a smartphone app. Those with symptoms 4 weeks or more after a positive polymerase chain reaction test for the virus, even if they were initially asymptomatic, were classified as having Long Covid. Among 41,361 vaccinated people infected between June and November 2021, when Delta was predominant, 10.8% reported Long Covid symptoms, such as fatigue, shortness of breath, or brain fog, the group reported in *The Lancet* last week. Of 56,003

infected between late December 2021 and early March, when Omicron was predominant, 4.5% reported symptoms more than 1 month later. But the Omicron surge in the United Kingdom was so large, peaking in late March at an estimated 350,000 symptomatic cases per day, that it helped drive large numbers of new Long Covid cases despite the reduction in risk.

U.S. backs shots for very young

COVID-19 | U.S. regulators have given a long-awaited thumbs up to start vaccinating nearly 20 million of the country's youngest children against COVID-19. The Centers for Disease Control and Prevention (CDC) on 18 June recommended all children 6 months through 5 years old be inoculated; a day earlier, the U.S. Food and Drug Administration (FDA) had acted on a unanimous 15 June recommendation by external advisers and authorized the Pfizer-BioNTech and Moderna vaccines for preschool children. The Moderna vaccine for children requires two doses; the Pfizer-BioNTech vaccine, three. FDA authorized the Pfizer vaccine for children 6 months to 4 years old; it was already authorized for 5-year-olds. The agency authorized the Moderna shot for children 6 months to 17 years old. CDC data through April list COVID-19 as the fifth leading cause of death in children 1 to 4 years old and the fourth leading cause in infants. The United States will become the first country to give COVID-19 vaccines to children as young as 6 months old, according to a White House statement.

Teams tackle cancer challenges

FUNDING | Four multi-institutional teams will receive $25 million each over 5 years to explore vexing questions in cancer after their research proposals were selected for funding from the Cancer Grand Challenges, a U.K.-U.S. collaboration. The charity Cancer Research UK started a similar program with awards in 2017 and 2019, then partnered with the U.S. National Cancer Institute for the latest round, announced last week. The multidisciplinary teams led by U.S. and European researchers will study using immune cells to treat solid tumors in children; the biology of DNA found outside a cancer cell's chromosomes; how cells that develop mutations turn cancerous; and cachexia, a muscle wasting disease common in cancer patients.

Ex-professor gets 1 year in prison

RESEARCH SECURITY | Simon Ang, a former engineering professor at the University of Arkansas, Fayetteville, was sentenced last week to 366 days in prison for lying to the FBI about his status as an inventor. Ang is one of two dozen academic scientists who have been prosecuted under the government's 3-year-old China Initiative. In January he pleaded guilty to the felony charge after the U.S. government agreed to drop its allegations that he had hidden ties to China on federal grant applications. His sentence, which could be shortened, begins on 20 July. Ang was also fined $5500.

Web conferencing spurs novelty

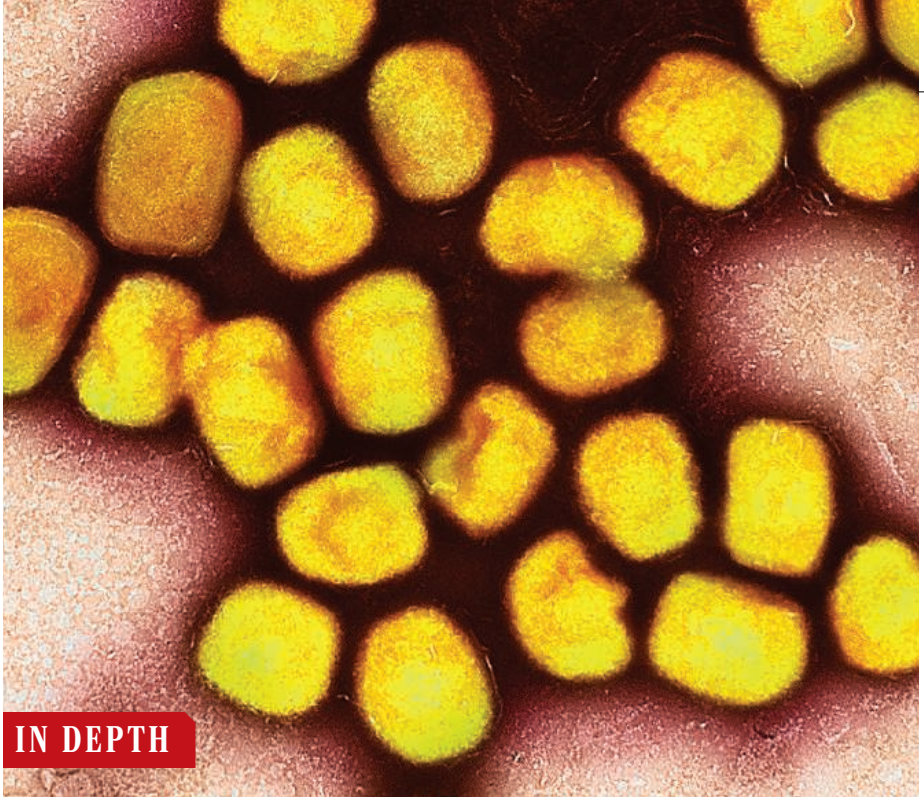
INNOVATION | Research teams have produced a larger number of innovative ideas even as they spread out across countries and institutions, and the progress may be driven by online collaboration platforms such as Zoom, a study has found. The share of all teams that are geographically dispersed has grown for decades. But until 2010, analysts recorded a decline in disruptive new ideas from dispersed researchers, likely because they pursued more incremental research and their scattering caused members to "silo" and sapped their creativity. But the slide has reversed since Slack, Microsoft Teams, and other platforms debuted in the past decade, researchers from the University of Oxford report in a recent working paper that analyzes journal articles by more than 10 million research teams in 11 academic fields from 1961 to 2020. (Teams that changed members were counted more than once.) Although in-person chats remain important for scientific creativity, an embrace of online tools has complemented those interactions, the authors say, and the COVID-19 pandemic may have accelerated this trend.

Drug for Alzheimer's amyloid fails

DRUG DEVELOPMENT | A decadelong clinical trial aiming to prevent cognitive decline in a Colombian family genetically predisposed to Alzheimer's disease has ended in failure, its sponsor, Roche, announced last week. The results are another blow to hopes that drugs removing the sticky brain protein beta-amyloid before symptoms emerge might slow the disease. The trial tested a drug called crenezumab, which—like many other amyloid-targeting antibodies—has already failed to show benefits in large trials of people with mild, early signs of disease. Researchers enrolled 252 extended family members, two-thirds of whom have a mutation, in a gene called *PSEN1*, that typically causes people to develop Alzheimer's symptoms in their mid-40s. Crenezumab didn't prevent or slow cognitive decline compared with a placebo, Roche subsidiary Genentech announced in a 15 June press release.

IN FOCUS Tiny cells meet large planets: Sarah Adkins-Jablonsky created an image of the Solar System by using ink and growing different colored strains of soil bacteria, including the antibiotic producer *Streptomyces*, on a growth medium. The title, *Ode to Kate Rubins*, honors a biologist who was the first astronaut to sequence DNA in space. The work was displayed last week at the annual meeting of the American Society for Microbiology.

IMAGE: AMERICAN SOCIETY FOR MICROBIOLOGY/SARAH ADKINS-JABLONSKY



A colorized transmission electron micrograph of monkeypox virus particles. The virus has sickened more than 3000 people in more than 30 countries in the past few months.

IN DEPTH

INFECTIOUS DISEASE

Why monkeypox is mostly hitting men who have sex with men

The once slow spreading virus may have found a new niche in tightly connected sexual networks

By **Kai Kupferschmidt**

Ever since monkeypox started to sicken thousands of people worldwide this spring, two big questions have loomed: Why is a virus that has never managed to spread beyond a few cases outside Africa suddenly causing such a big, global outbreak? And why are the overwhelming majority of those affected men who have sex with men (MSM)?

A long history of work on sexually transmitted infections (STIs) and early studies of the current outbreak suggest the answers may be linked: The virus may have made its way into highly interconnected sexual networks within the MSM community, where it can spread in ways that it cannot in the general population.

An epidemiological modeling study posted as a preprint last week by researchers at the London School of Hygiene & Tropical Medicine (LSHTM) supports that idea. It suggests the outbreak will keep growing rapidly if it isn't curtailed. The model also has implications for how to protect those most at risk and limit spread, while suggesting the risk for the wider population remains low.

But there are still many uncertainties, and communication is fraught because of the risk of stigmatizing MSM—and because communicating frankly about sexual behavior is hard. "I think we have to talk more about sex," says Yale School of Public Health epidemiologist and former HIV activist Gregg Gonsalves. "Everybody has been very clear about stigma, and saying it over and over again. The point is that you still have to address the risk of infection in our community."

Since early May, more than 3000 monkeypox cases have been reported in more than 30 countries where the virus is not normally seen. (Outbreaks are more common in many countries in West and Central Africa, where the virus has animal reservoirs. More than 60 cases and one death have been confirmed there this year.)

The vast majority of cases in the current outbreak have been in MSM. Researchers at the UK Health Security Agency (UKHSA), for example, asked patients to fill out questionnaires. Of 152 who did, 151 said they were MSM, the team wrote in a technical briefing published on 10 June; the remaining patient refused to answer. Other countries have seen similar patterns.

That could be a skewed picture, of course. "MSM have a better relationship with medical practitioners than heterosexual men," says Lilith Whittles, an infectious disease modeler at Imperial College London, which could mean they are more likely to report monkeypox symptoms and get tested for the virus. "I don't know that we're necessarily looking enough in heterosexual social networks to make the conclusion that this is not a broader problem," says Boghuma Titanji, a virologist at Emory University who works at a sexual health clinic.

But most researchers say such "ascertainment bias" is unlikely to explain the striking pattern. Although some monkeypox patients have mild infections that may be missed or misdiagnosed, others have very characteristic rashes and agonizing pains that require hospitalization for pain treatment. If many people outside the MSM community had monkeypox, more would have shown up in the statistics by now.

Ashleigh Tuite, an infectious disease epidemiologist at the University of Toronto, says she "understands the hesitation" to focus on MSM, given the risk of stigma that could worsen discrimination and cause those who are affected to delay seeking care. "But based on the data that we have, and based on the contact tracing that's been done, it's very clear that this is an MSM-focused outbreak at this point," she says. "Anyone can get monkeypox, but we're seeing disease activity primarily among" MSM, confirms Demetre Daskalakis, an HIV prevention specialist at the U.S. Centers for Disease Control and Prevention.

Sexual encounters clearly play a role in transmission. Of the 152 people in the UKHSA data set, 82 were invited for additional interviews focusing on their sexual health. Among the 45 who participated, 44% reported more than 10 sexual partners in the previous 3 months, and 44% reported group sex during the incubation period. Exactly how the virus is passed on is less clear. Researchers have found viral DNA, and even infectious virus, in the semen of some patients, but they aren't sure that is important for transmission; skin-to-skin contact may be enough. (Other STIs, including herpes and scabies, also primarily spread this way.)

To those who study how pathogens spread through social and sexual networks, the pattern is not a big surprise. In the 1970s

PHOTO: NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DIS

and '80s, researchers trying to understand the spread of sexually transmitted diseases such as gonorrhea were stumped, says LSHTM epidemiologist Adam Kucharski: Survey data showed people's average numbers of sexual partners were too low to sustain transmission. But averages obscured that although many people have few sexual partners, some have large numbers. This helps drive transmission because, if infected, they are more likely to infect others.

Sexual networks among MSM are not different in nature from those of other groups, Whittles stresses, but a core group of people are much more densely connected than in other communities. They change partners more frequently and are more likely to have several partners at the same time. "These things occur in all sexual networks, it is just a question of the degree," Whittles says. And in a densely connected network, the virus is less likely to hit a dead end.

"It's entirely possible for this epidemic to rage among a subset of people just because that subset is connected in a network differently than everyone else," says Keletso Makofane, a social network epidemiologist at the FXB Center for Health and Human Rights at Harvard University. Makofane hopes to launch a study in New York City in August to better understand the spread of the disease. "The idea is to get a sense of how many people report symptoms that are consistent with monkeypox and how they are connected," he says.

"I think we have to talk more about sex."
Gregg Gonsalves, Yale School of Public Health

The LSHTM study, posted on medRxiv on 13 June, used U.K. data about sexual partnership patterns to model the spread of monkeypox among MSM and outside that group. Because it is not yet clear how contagious the virus is, the researchers modeled scenarios based on different risk levels. Without effective intervention measures or behavioral changes, a large and sustained outbreak with more than 10,000 cases among MSM globally is "highly likely," they write. "In contrast, sustained transmission in the non-MSM population is unlikely in all scenarios considered."

Because the model is based on U.K. data, the findings may not apply elsewhere, first author Akira Endo says. And other factors may have made the outbreak worse. Monkeypox may have mutated in ways that allow it to transmit more easily, and the share of the population who have had the smallpox vaccine—which also offers some protection against monkeypox—is declining because smallpox vaccination was abandoned worldwide starting in the 1970s. But the modeling shows "we do not necessarily need [those factors] to explain the observed patterns," Endo says.

The modelers' conclusions put epidemiologists in a delicate position, and some declined to talk to *Science* for fear of stigmatizing MSM. Endo says he understands that and agrees the findings could easily be misunderstood. "Meanwhile I also understand that there's a risk in the other direction—that the information doesn't reach those who need it most before it's too late."

Whittles agrees, calling the findings "practical information, in terms of where it's spreading. It is a morally neutral thing," she says. "Knowledge of what's happening is power, even though that knowledge is imperfect and will change," Daskalakis adds.

The virus could still find other nonsexual networks with similar characteristics. Daskalakis recalls a U.S. outbreak of methicillin-resistant *Staphylococcus aureus* in the 2000s that started in the MSM community but later spread in gyms, among athletes, and in prisons. Monkeypox could also start to spread among sex workers and their clients, Tuite says.

How fast the virus will spread in the months ahead depends on control efforts. National health authorities in Europe, Canada, and the United States have issued guidance about how to reduce the risk of infection, and dating apps have alerted users to the risk of monkeypox and to its symptoms, which may change contact patterns. Increasing awareness among health care workers may have an impact as well, Whittles says: Faster diagnoses mean patients will isolate sooner in their infection, reducing onward transmission. "So there's a couple of different ways that behavior can change, even if it's not people having less sex," she says. And as infected individuals develop immunity, "we may see a slowdown of the outbreak earlier than we might imagine," Endo says.

Many countries are gearing up to start immunization campaigns. Targeting those with many sexual partners may be the most efficient approach. In guidance published on 21 June, UKHSA announced it would start to offer vaccines to MSM at highest risk. "Risk criteria would include a recent history of multiple partners, participating in group sex, attending 'sex on premises' venues or a proxy marker such as recent bacterial STI," the agency said.

It's important to alert that community and do it the right way, Gonsalves says. "We should say: It's not about who you are. It's about what you're doing. And we're not going to stigmatize it. But just know that you're at greater risk if you fit this profile." ■

NEUROSCIENCE

Stimulating the brain may help people who stutter

Trials suggest small electric currents aid fluency, but how long the benefits last is unclear

By **Ignacio Amigo**

When Guillermo Mejias was 7 years old, his parents sent him out to buy bread during a family holiday in southern Spain. Mejias still remembers his growing anxiety as he walked to the bakery, repeating what he would say over and over in his head. But when the moment arrived, he was unable to produce a single word. He recalls returning empty-handed, ashamed, and wondering what to tell his parents. "I was so tense that I had been inadvertently biting my cheeks and tongue and my mouth was bleeding," he says.

Mejias still stutters, but today, as a brain researcher at the Complutense University of Madrid, he investigates ways to treat the problem. He is part of a growing group of researchers who have pinned their hopes on noninvasive brain stimulation, a set of techniques that applies small electric currents to specific brain regions. A few years ago, Mejias tried such a technique on himself in a one-off experiment that temporarily reduced the frequency of his stuttering, he says. A few randomized trials, including one published this month in the *Journal of Fluency Disorders*, also suggest brain stimulation can benefit people who stutter. "I think brain stimulation is the future," Mejias says.

Not everyone agrees. No study has yet reported effects lasting months after treatment, let alone years. And anecdotal evidence suggests trials of stuttering treatments are susceptible to the placebo effect, meaning that, just from being in the trial, participants may see positive effects that aren't due to the treatment.

About 1% of the world's adult population—some 70 million people—is thought to stut-

Julio Prieto (left) and a colleague treat a person who stutters with noninvasive brain stimulation.

PHOTO: GREGORIO MARAÑÓN UNIVERSITY GENERAL HOSPITAL

ter, which can impair quality of life and cause social distress and stigma. But its causes are still poorly understood.

The only available treatment, speech therapy, can improve fluency to some extent and help avoid side effects of stuttering, such as involuntary movements and tics. But its success rate is limited in adults and it fails to address the root of the problem, Mejias says: "Treating stuttering with speech therapy is akin to treating Parkinson's disease with physiotherapy."

Noninvasive brain stimulation, however, directly stimulates or inhibits neurons in specific brain regions in an effort to remodel circuits of interconnected cells. It has already shown benefits for depression—particularly in people who are resistant to drug treatments—and in stroke rehabilitation, where it can help improve movement and speech.

Mejias and others test its effects in stuttering by using various techniques to apply electrical stimulation to circuits involved in speech. Next, participants might perform a task that activates those same regions again, such as reading a text in unison with another person or with a metronome, which many people who stutter can do with some degree of fluency.

The hypothesis is that performing these tasks after the stimulation can help reshape the troublesome circuits. Although the electrical currents can cause slight discomfort, noninvasive brain stimulation is widely considered safe.

For his Ph.D., Mejias investigated the effects of a technique called transcranial magnetic stimulation (TMS), in which a magnetic field is applied through a small coil placed on the head of the participant, inducing a small electric current in the brain. The coil is only a few millimeters wide, allowing the stimulation to be targeted to precise brain areas. Mejias focused on the supplementary motor area, a region of the cerebral cortex involved in movement control that shows abnormal functioning in people who stutter.

Because the study was largely exploratory—and to avoid the red tape of a larger trial—Mejias decided to test TMS on himself. In a short paper published in 2019, he reported that the treatment reduced the blocks and repetitions in his speech by 30%. "I felt a progressive improvement, which also helped me build up my confidence," he says. "The effects were evident for a few weeks but we didn't monitor them consistently so I don't know exactly how long they lasted."

Kate Watkins, a neuropsychologist at the University of Oxford, was the first to treat people who stutter with another technique called transcranial direct current stimulation (tDCS), which passes an electrical current from one side of the brain to the other. Because the electrodes used in tDCS are relatively large, researchers have less control over where the stimulation happens than with TMS, but the required equipment is much simpler and cheaper. "If it works, it's something that speech and language pathologists can use very easily and safely," Watkins says. Another benefit: It often causes less discomfort than TMS.

In a randomized, double-blind trial with 30 participants published in 2018, Watkins and colleagues applied direct currents during 20 minutes for five consecutive days. They aimed to stimulate the left inferior frontal cortex, a brain region involved in language processing and speech that shows reduced activity in people who stutter. To control for the placebo effect, half of the participants received a sham stimulation with a minuscule current—enough for them to feel something, but too weak to induce an effect.

The benefit was statistically significant, but modest: After 1 week, participants who received the brain stimulation stuttered 8% of their syllables, down from 11% at baseline, with no detectable changes in the control group. The study published this month—by researchers in Iran, Germany, and Canada—involved 50 participants and was similar in both design and outcome, with stuttered syllables going down from 8.5% to 5.4% after stimulation.

"The results of improvement are evident and similar in both," says Julio Prieto, chief of the Clinical Neurophysiology Service of the Gregorio Marañón University General Hospital in Madrid, who also studies stuttering and has collaborated with Mejias. "From my point of view they both have the same limitations," he adds: few treatment sessions and a short follow-up period, making it impossible to say whether effects last.

Prieto says the field would benefit from more standardized protocols. Like many noninvasive brain stimulation studies, stuttering experiments vary in design and equipment. Trials also target different brain regions and use currents of different intensity and duration. "And then there are also factors that you can't control, such as sex differences, the particpant's alertness, and even genetic variations," Prieto says. Several studies suggest there's a large variability in responses to brain stimulation between individuals.

Dennis Drayna, an emeritus geneticist at the U.S. National Institutes of Health who works on stuttering, says skepticism is in order. "Dozens of different treatments for stuttering have been tried, often with spectacular initial success," he says. "None of these have lasted."

Brain stimulation researchers hope to persuade the skeptics. Mejias is seeking funding to expand his self-experiment into a larger study. Watkins and her team are planning to try yet another type of electrical stimulation called transcranial alternating current stimulation, in which the current's amplitude is not fixed but oscillates. "It mimics the natural brain rhythms," says Birtan Demirel, a researcher at Watkins's laboratory who also stutters.

Mejias is optimistic about treating stuttering at its roots. "The brain used to be something almost mythical, nobody knew how it worked. Now, we're beginning to understand it and in the future we'll be able to modify it." ■

Ignacio Amigo is a journalist in Madrid.

SCIENCE POLICY

Arati Prabhakar to lead Biden's science office

Applied physicist would bring wealth of policy experience as successor to Eric Lander

By **Jeffrey Mervis**

In 1993, then-President Bill Clinton picked a 34-year-old applied physicist named Arati Prabhakar to lead the National Institute of Standards and Technology (NIST). Two decades later, then-President Barack Obama chose her to lead the Defense Advanced Research Projects Agency (DARPA). And this week, President Joe Biden nominated her to be director of the White House Office of Science and Technology Policy (OSTP) and made her his science-adviser-in-waiting.

If confirmed by the Senate, the 63-year-old Prabhakar will succeed Eric Lander, who resigned on 7 February after admitting to bullying his staff and creating a hostile work environment during his 9 months in office. The twin jobs would give her a role in addressing thorny science policy issues including how best to position the United States to compete with China, implement workable rules for protecting U.S.-funded academic research from theft, and reduce inequality in the research community. Biden said in a 21 June statement that Prabhakar would become his science adviser upon confirmation as OSTP director.

Prabhakar's extensive experience in Washington, D.C., and technical savvy will be a huge plus, say those who know her.

"I have found Arati to be very smart, very principled, and … with excellent leadership qualities," says John Holdren, who led OSTP for 8 years and served as Obama's science adviser. Her reputation as a team player is an asset as well, adds Washington, D.C., lobbyist Bart Gordon, a one-time chair of the House of Representatives science committee as a Democratic representative from Tennessee. "She has all the background you could ask for, and she's also such a nice person," says Gordon, now with K&L Gates.

Sudip Parikh, CEO of AAAS (which publishes *Science*), thinks Prabhakar will "bring an important voice to equity issues across the scientific enterprise" as the first woman and first person of color to lead OSTP and serve as science adviser. Biden's announcement notes that she would also be the first immigrant to hold those jobs.

Born in India and raised in Texas, Prabhakar earned her Ph.D. from the California Institute of Technology in 1984 and then went to work for the government. She spent 7 years as a program manager at DARPA, the military's technology incubator, before becoming the first woman to lead NIST.

In 1997, she moved to the West Coast, where she spent more than a decade as a venture capitalist in Silicon Valley. In 2012, she became the second woman to lead DARPA, serving a 5-year stint in which she created a biotechnology office that pioneered work on RNA vaccines, key to fighting the current pandemic. In 2019, she formed Actuate, a nonprofit that works with philanthropy to conduct what she calls "solutions R&D" in areas ranging from sustainable energy and health to the ethical use of technology. After Biden was elected, some policy wonks bet he would name Prabhakar as his science adviser.

Arati Prabhakar, shown in 2015, has led several technology agencies.

PHOTO: MANDEL NGAN/AFP/GETTY IMAGES

The adviser's main job is to help carry out the president's agenda for science, which Biden described in a 15 January 2021 letter after nominating Lander. It includes applying lessons from the pandemic to improve public health, enlisting research to combat climate change, ensuring that the country remains a leader in emerging high-tech fields, reducing inequality in the research community, and turning federally funded basic research into jobs and products.

Congress has helped Biden achieve a few of those goals, passing legislation that creates a new Advanced Research Projects Agency for Health and a new technology directorate at the National Science Foundation. But Biden's $2 trillion social and economic recovery plan, which includes significant investments in sustainable energy, is all but dead because it lacks support from Senator Joe Manchin (D–WV), the deciding vote in an evenly split Senate. And legislators are still negotiating a massive research, manufacturing, and trade bill designed to strengthen the country's ability to compete with China.

If Republicans regain control of one or both chambers of Congress in this fall's elections, the Biden White House will likely need to rely on executive orders to implement its agenda, rather than on any new legislation.

Lander, a mathematician–turned–molecular-biologist, worked under Obama with then-Vice President Biden on the Cancer Moonshot and co-chaired the President's Council of Advisors on Science and Technology. Prabhakar doesn't have that personal link. But her technology background will be an asset, predicts William Bonvillian, a former Senate science staffer and federal relations director for the Massachusetts Institute of Technology. "We're in a huge fight with China over technology, and Arati is steeped in all aspects of that issue," he says. "She understands the role of technology in the defense sector and how to responsibly manage competition with China, and she's worked with the private sector on high-tech startups."

As a midterm appointment, Prabhakar probably won't spend much time crafting new initiatives. But she'll still have plenty to do. "OSTP has influence, not power," one longtime Washington, D.C., insider says about the office, which draws heavily on staff seconded from other federal agencies. "And it has as much influence as the director and her staff can muster." ■

A sex worker in Beira, Mozambique, gets tested for HIV. An injectable prophylactic drug can give people at high risk a new way to prevent infection.

PHOTO: GIANLUIGI GUERCIA/AFP/GETTY IMAGES

GLOBAL HEALTH

HIV prevention tool is too slow to reach Africa, activists say

A long-lasting injectable drug was tested in seven African countries. But it's unclear when they'll have access to it

By **Andrew Green**, *in Kampala, Uganda*

Melb Simiyu, an HIV prevention officer at a support organization for sex workers here, says most of her clients have asked when a drug called CAB-LA will become available. Approved by the U.S. Food and Drug Administration (FDA) in December 2021, the HIV-prevention drug could drastically reduce infections among marginalized groups like the one she works with.

But even though a trial in Uganda and six other sub-Saharan African countries provided data key for FDA's approval, Africa, still the part of the world hardest hit by HIV, may face a long wait to get the drug at an affordable price.

On 27 May, CAB-LA's developer, Viiv Healthcare, made a formal commitment to issue a voluntary license for the drug, an injectable form of pre-exposure prophylaxis (PrEP), to the Geneva-based Medicines Patent Pool (MPP). If the two reach an agreement, MPP could broker deals with generic manufacturers to produce cheap versions of CAB-LA for low- and middle-income countries. But it may take years before those generics become available, and activists accuse Viiv of dragging its feet, which the company denies. Zambian HIV activist Kenly Sikwese, who coordinates the African Community Advisory Board, says equitable access in sub-Saharan Africa should have been guaranteed before people from the region were recruited for clinical trials. "Viiv's got to be serious about making the product accessible to everyone who needs it," Sikwese says.

Two oral versions of PrEP are already available, both as a daily pill. They have been shown to reduce HIV infection in groups at high risk, including sex workers and men who have sex with men (MSM), by up to 99%—but only when taken consistently, which many users find hard. CAB-LA promises to be more effective because adherence is much easier. The key ingredient, an antiviral compound named cabotegravir that prevents the viral genome from integrating into the DNA of host cells, is injected into the buttock and slowly released into the body. The bimonthly injections mean less hassle and more discretion for the user than a daily pill.

The trial in sub-Saharan Africa, which enrolled 3200 women, showed CAB-LA to be 89% more effective at preventing new HIV infections than oral PrEP. It also proved superior in another study, in nearly 4600 MSM and transgender women across four continents. Matthew Kavanagh, deputy executive director of the Joint United Nations Programme on HIV/AIDS (UNAIDS), calls CAB-LA "potentially the most effective biomedical prevention tool to come along in a decade."

The World Health Organization is expected to issue safety guidelines for CAB-LA in July, which should spur countries to follow FDA's lead and approve it. UNAIDS has had initial conversations about financing its distribution with donors, Kavanagh says, including the U.S. President's Emergency Plan for AIDS Relief, the major funder of oral PrEP in Africa. There's also "a handful of governments that say they would like to invest in this," he says.

But all are waiting to find out how many vials will be available and at what cost. Viiv has priced injectable PrEP at $22,000 for a year's supply in the United States. For the world's poorest countries, that figure needs to come down to the current price of generic oral PrEP, less than $100 annually, Kavanagh says. "If it's a lot more than that, we're sunk," he says; that would make CAB-LA an "inequality booster."

Viiv has pledged to provide CAB-LA at nonprofit prices to the poorest countries, including all of sub-Saharan Africa. But the company will not say what that price is, and in March, it issued a vaguely worded statement suggesting a voluntary license would not help achieve the goal of affordable access. The move perplexed and enraged activists, and the statement has since been taken down. Harmony Garges, Viiv's chief medical officer, says there was a "misunderstanding about our commitment" and that the company was only "signposting that this was going to be different, more complex" than manufacturing generic oral treatment, for which Viiv has provided a voluntary license in the past.

The new commitment to work with MPP is "a step in the right direction," says Yvette

Raphael, executive director of Advocacy for Prevention of HIV and AIDS in South Africa. But, she adds, "It should have been done the minute the studies were stopped" in 2020, given that setting up generic CAB-LA production could take up to 3 years. Activists also note that it's unclear how many low- and middle-income countries will be allowed access to generics under a license and whether Viiv will share technology to help manufacturers ramp up production.

"I appreciate from the activist side that there's always a level of impatience," Garges says, but negotiations with MPP started just 6 months after CAB-LA received its first regulatory approval. "I don't view this as any kind of delayed approach."

Lowering the price isn't the only hurdle to providing the drug. HIV can develop resistance to cabotegravir if it is given to a person who is already infected, and the resistance can hamper other integrase inhibitors, including dolutegravir, the backbone of current HIV treatments. That's why FDA requires injectable PrEP recipients in the United States to first undergo a sensitive HIV test. But those cost money and can be time-consuming to conduct, says Linda-Gail Bekker, director of the Desmond Tutu HIV Centre in Cape Town, South Africa. If regulators determine such tests are needed, officials will have to consider how to make them available, Bekker says.

Advocates are keen to avoid what happened with the global rollout of oral PrEP, which began in 2015 but was quickly ensnared in "demonstration projects that were too small and too slow," says Mitchell Warren, executive director of AVAC, an HIV prevention advocacy group. By 2020, fewer than 1 million people in developing countries were on PrEP, far short of the UNAIDS goal of 3 million.

AVAC identified several other missteps to avoid this time. Campaigns that framed oral PrEP as a prevention tool specifically for those at high risk of HIV infection made some people hesitant to ask for it for fear of stigma. Requiring people to visit busy clinics to start PrEP or pick up refills heightened those fears.

Taking that lesson to heart, Andrew Mujugira, a senior research scientist at Uganda's Infectious Diseases Institute, hopes to recruit and train health workers from within communities of sex workers, MSM, and trans people in Uganda to deliver CAB-LA. Whenever the drug arrives, he wants to be prepared, he says: "If it can be procured, if somebody can pay for it and it can be accessible, it's game changing." ■

Andrew Green is a journalist based in Berlin.

PHOTO: DMITRY KOVALCHUK/ISTOCK.COM

WORKFORCE

U.S. labs face severe postdoc shortage

Early-career researchers increasingly avoid low-paying, insecure postdoc positions, a *Science* investigation finds

By **Katie Langin**

When geneticist Jennifer Mason posted an ad for a postdoc position in early March, she was eager to have someone on board by April or May to tackle recently funded projects in her lab at Clemson University. Instead, it took 2 months to receive a single application. Since then, only two more have come in. "Money is just sitting there that isn't being used … and there's these projects that aren't moving anywhere as a result," Mason says.

She isn't alone. On social media, many U.S. academics have been pointing to widespread challenges in recruiting for postdoc positions, which typically offer a few years of temporary employment to researchers who recently finished their Ph.D. To examine the extent of the problem, *Science* contacted more than 100 U.S.-based researchers who advertised for postdoc positions this year on scientific society job boards; of the 37 who responded with information about their hiring experiences, three-quarters reported challenges recruiting. "This year is hard for me to wrestle with: … we received absolutely zero response from our posting," one wrote. "The number of applications is 10 times less than 2018-2019," said another.

Those experiencing challenges span STEM fields, including biomedicine, chemistry, environmental science, anthropology, physics, and computer science. Many reported a drop not only in the total number of applications, but also in the quality of applications. "It took two rounds of advertising my current postdoc opening—once in October 2021 and again in April 2022—to find a competitive applicant," one researcher wrote by email. "To find qualified people, it's way more difficult than it used to [be]," says Donna Zhang, a professor of pharmacology and toxicology at the University of Arizona who is trying to hire multiple postdocs. "It's very bad."

It's also not what some predicted 2 years ago when the pandemic hit and faculty job

openings dried up. At the time, the fear was that postdocs would stay in their positions longer, leaving few openings for new Ph.D. graduates. But the faculty job market rebounded in 2021, according to a preprint posted on bioRxiv last month, enabling some postdocs to move on from their temporary positions. And the wider labor market has seen dramatic changes because of what some are calling "the Great Resignation." "There are jobs everywhere," says Donna Ginther, a professor of economics at the University of Kansas, Lawrence, who studies the scientific workforce. Postdocs in general aren't paid well, she points out, and amid the current labor shortage, higher paid jobs outside academia have become more available. "Ph.D.s are looking at the labor market, seeing opportunities out there, and taking them," she says.

The shift is disruptive for academic researchers looking to staff their labs, acknowledges Cynthia Fuhrmann, an associate professor at the University of Massachusetts Chan Medical School who serves as a principal investigator (PI) for the Professional Development Hub. But she is pleased to see Ph.D. graduates "voting with their feet" and finding ways to contribute outside the Ivory Tower. "They recognize that there's exciting career opportunities out there that don't actually necessitate postdoc training," she says.

> **"Ph.D.s are looking at the labor market, seeing opportunities out there, and taking them."**
> **Donna Ginther**, University of Kansas, Lawrence

Graduate students are also listening to current postdocs, who in recent years have become increasingly vocal about the precarious nature of their positions and the challenges of getting by on a postdoc salary, especially in areas with a high cost of living, says Ariangela Kozik, a microbiologist and research investigator at the University of Michigan, Ann Arbor, who led the study of faculty job openings. During her postdoc, which wrapped up last year, she and her husband struggled to afford child care, for instance. "It's an investment [to do a postdoc] and you take it at a loss," she says. "And I think the more that people are aware of that, fewer and fewer people are going to be willing to take that risk—especially because there's no guarantee of a tenure-track position at the end of it." For early-career researchers who aren't economically privileged, it's an especially tough road, she says. "Preexisting privilege … determines whether or not you're able to stick it out" through the economic hardships of a postdoc position, she says. "It shouldn't be that way."

Many faculty members acknowledge academic postdocs should be paid more. But they say that often, their hands are tied. Postdoc salaries are frequently based on what the U.S. National Institutes of Health sets as its standard, "and that's pretty low," says Daniel Wolf Savin, a physicist and senior research scientist at Columbia University who is currently struggling to fill five postdoc positions. When hiring, he competes with national labs that offer up to $20,000 more per year. If he were to propose to match those salaries in a grant application, "the program office is going to look at it and say, 'Look, I can't give you this much money. It's so out of line with what everyone else is asking for,'" he says.

"The PIs—I feel for them [because] they can't solve the problem by themselves," Kozik says. Many of the policies that determine postdoc salary and working conditions are set by funding agencies and universities, she adds.

Others say many young researchers became disenchanted with academia during the pandemic. "A lot of universities really showed their cards about whether they care about people," says Gary McDowell, CEO and founder of Lightoller consultancy and an advocate for early-career researchers. For instance, some universities forced academics to work on campus and continued to expect high productivity despite COVID-19 exposure risk and ongoing disruptions, he says.

McDowell thinks the apparent shortage of talented early-career researchers willing to sign on as postdocs will serve as a wake-up call for universities. "I honestly think that's the only way that academia will change and start to actually address these problems," he says.

Fuhrmann agrees. "This could be a really great opportunity to see these pressures on the system and think about intentionally creating systemic change," she says. For instance, academic administrators and policymakers may want to ask, "How can we make postdoc positions more attractive?" They may also want to address the career insecurity that comes with postdoc positions, she adds. "How can we diversify the academic workforce so that perhaps we have more research scientist positions, as well as postdoc positions … so that students see additional paths—pathways for promotion after their postdoc training?"

As Kozin puts it, "It really is going to need to take an all-stakeholder effort here to turn some things around." ■

SCIENTIFIC MISCONDUCT

Disgraced stem cell surgeon convicted of criminal harm

Swedish court gives Paolo Macchiarini suspended sentence for death of windpipe transplant patient

By **Gretchen Vogel**

A surgeon who just a decade ago was celebrated around the globe as a pioneer in stem cell transplants has been convicted of one count of "causing bodily harm," a felony, in a Swedish court. The district court in Solna on 16 June found Paolo Macchiarini not guilty on other charges, including aggravated assault relating to three patients he treated while working for Sweden's famed Karolinska Institute (KI). The court said the penalty was "a suspended sentence," but did not specify how long the sentence would be if imposed. Prosecutors had asked for a 5-year prison sentence.

The verdict is the latest development in Macchiarini's stunning fall from grace. In 2010, the Italian surgeon was recruited by KI—home of the committee that awards the Nobel Prize in Physiology or Medicine. A year later, he began to treat patients whose windpipes were damaged by injury or disease by implanting synthetic versions, seeded with stem cells isolated from the patients' own bone marrow. He claimed the cells would grow and integrate with the patients' tissue. The operations were hailed at the time as a breakthrough in regenerative medicine, although some observers remained skeptical. Macchiarini was fired in 2016 amid allegations of fraud and scientific misconduct after many of his transplant recipients died.

Macchiarini has consistently claimed he is innocent. In court testimony, he defended himself as only trying to help desperately ill patients, and said he had the full support of KI and his colleagues. His lawyer did not respond to emails from *Science*.

The mild sentence comes as a disappointment to Macchiarini's critics. Cardio-

thoracic surgeon Matthias Corbascio, one of Macchiarini's former colleagues at KI who first raised questions about the work, describes it as "terrible" and "insane." "The court basically gave him a slap on the wrist," says Corbascio, who is now at the University of Copenhagen. "There are people who go to jail for 5 years for not paying their taxes. This guy mutilated people."

Between 2011 and 2014, Macchiarini implanted artificial windpipes in at least eight patients in Sweden, the United States, and Russia. All but one died following severe complications with the implants. (The patient who survived had the implant removed.)

Four of Macchiarini's colleagues at KI, including Corbascio, filed official complaints in 2014 charging that several papers describing the surgeries deliberately left out serious complications in the patients. They also questioned whether Macchiarini had obtained proper ethical permission for the surgeries and challenged a paper describing animal experiments with the technique.

KI commissioned surgeon Bengt Gerdin, professor emeritus at Uppsala University, to investigate the issue. Gerdin's report, released in May 2015, concluded Macchiarini was guilty of misconduct, but university administrators rejected that conclusion 3 months later, arguing Macchinarini and his co-authors had "satisfactorily countered" the issues Gerdin raised. They allowed Macchiarini to continue his work as a visiting professor with a research lab.

However, Swedish TV reporters also decided to take a close look at Macchiarini's work. Their devastating documentary, which aired in January 2016, led KI officials to reopen the investigation. Around the same time, an article in *Vanity Fair* described how Macchiarini had misled a girlfriend—an NBC news producer working on a flattering documentary about his surgeries—into thinking they would be married in a ceremony attended by the Clintons and Obamas and officiated by the pope. The story portrayed Macchiarini, who was already married at the time, as a serial fabulist who exaggerated or lied about degrees, academic appointments, and personal accomplishments.

KI fired Macchiarini in March 2016, 2 months after the Swedish documentary came out. In 2018, KI finally released the findings of its second investigation, which confirmed he had committed misconduct.

In 2017, prosecutors in Sweden charged Macchiarini with manslaughter in connection with the deaths of three patients who had received transplants at KI in 2011 and 2012: Andemariam Teklesenbet Beyene, a graduate student from Eritrea with a slow-growing cancer obstructing his windpipe; Christopher Lyles, a 30-year-old American with tracheal cancer; and Yesim Cetir, a teenager from Turkey whose trachea had been accidentally damaged during a previous surgery. Prosecutors dropped the case a few months later, however, saying they didn't have sufficient evidence to prove manslaughter.

They reopened the case in 2020, this time charging Macchiarini on lesser counts of aggravated assault and severe bodily harm. The court heard testimony in the case between 27 April and 12 May.

At the trial, Macchiarini and his lawyer argued that the patients, with no other options, had been treated under compassionate care standards. Other witnesses contradicted that, arguing that with Andemariam's slow-growing cancer, minor surgeries and other treatments could have kept him alive. "Every patient had another, nonlethal option," Pierre Delaere, a trachea expert at KU Leuven and an early critic of Macchiarini, said in an email to *Science*.

Paolo Macchiarini, in a 2013 photo, implanted synthetic windpipes into multiple patients; all but one died following complications with the implants.

PHOTO: LINA ALRIKSSON/DN/TT NEWS AGENCY

A panel of judges unanimously ruled that "the interventions were not in accordance with science and proven experience," according to a statement released last week. But the judges decided that, because the initial prognoses for Andemariam and Lyles were so grim without surgery, Macchiarini could not be held criminally responsible in those cases. In the case of Cetir, he should have known the risk of complications was higher than the expected benefit, the judges ruled. But they found prosecutors didn't prove Macchiarini was "indifferent" to the injuries or suffering that the surgeries might cause, the standard required for aggravated assault. Instead, they found him guilty of negligence and the lesser charge of causing bodily harm.

In Sweden, both the defense and the prosecution can appeal rulings, and senior prosecutor Karin Lundström-Kron says she and her colleagues will decide in the coming weeks. They will especially examine the judges' decision that Macchiarini couldn't be held responsible for what happened to Andemariam and Lyles "because he acted in distress," according to the court's statement. "That's something that was new to us, so we need to analyze that," Lundström-Kron says. She says it may mean the judges thought he and the patients believed time was short and options were few. Macchiarini's lawyer told reporters last week that his client is considering an appeal.

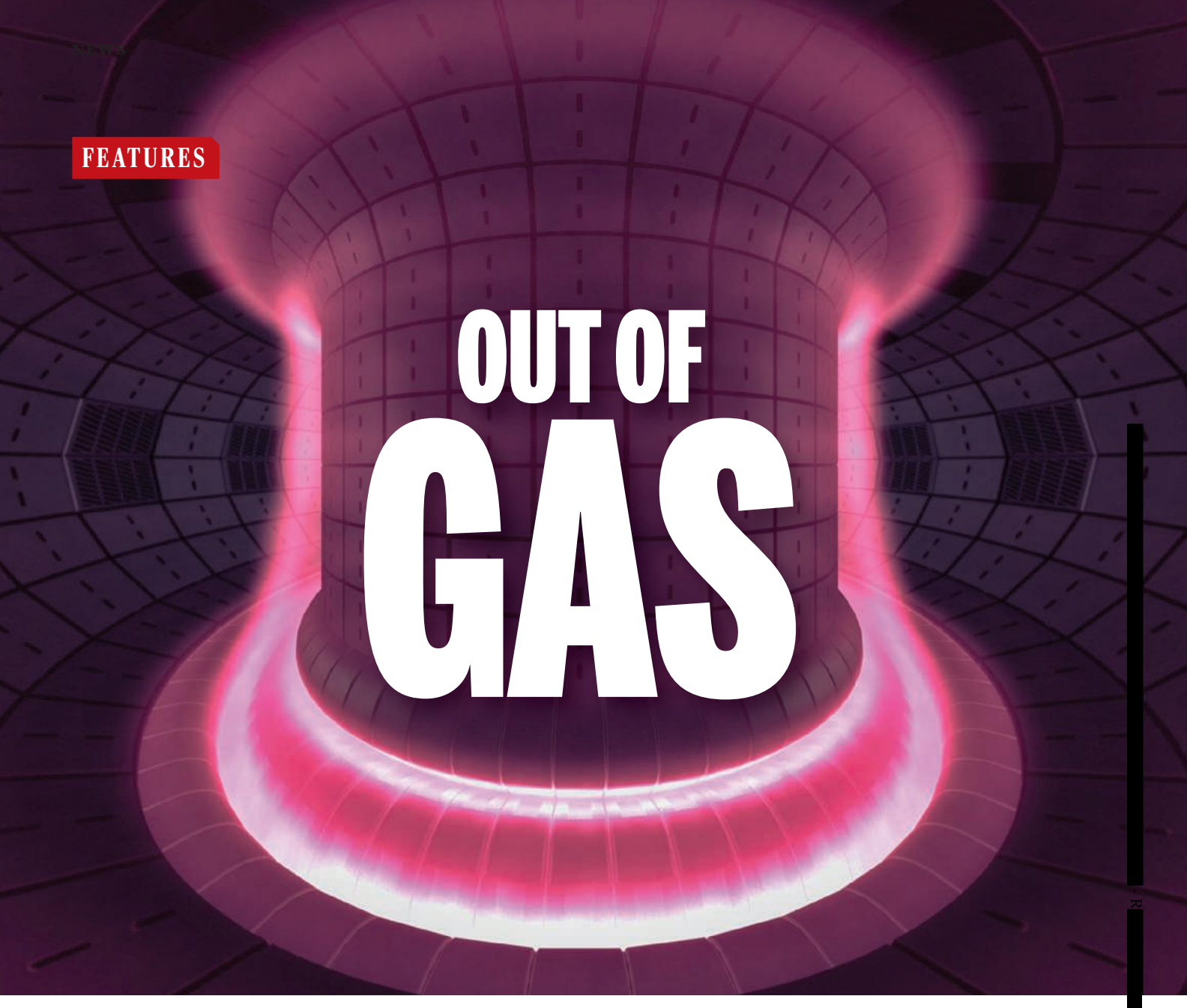
Delaere feels the sentence is too light for what he considers a serious crime. He says prosecutors should have pressed for manslaughter charges because Macchiarini had no evidence that the tracheas he implanted would develop into normally functioning organs. "When you implant a synthetic trachea you have one certainty, and that is that the patient will die as a result of it," Delaere says in the email. "Acquittal is therefore difficult to defend."

But Gerdin says the verdict is not unexpected. The court had to apply the "intentional indifference" standard for the assault charges, he says, which is difficult to prove. Macchiarini's argument that KI fully supported and encouraged the surgeries likely swayed the judges as well, he says. KI officials "let him do this. It relieved a little bit of burden from him."

Gerdin says the conviction on one count still shows "that health care doesn't stand above the law. If as a doctor you do stupid things, then you can be charged." But the not-guilty verdicts on the other two cases highlight how the law can leave vulnerable people unprotected, he says, for example when they are facing life-and-death decisions and holding less information than their doctor.

The felony conviction means Macchiarini would have a hard time finding employment again in Sweden, Gerdin says. But because Macchiarini is apparently living in Spain, it's unlikely to have much effect. "It's actually meaningless for him," Gerdin says. "But he's lost his reputation, and that's more important." ■

FEATURES



OUT OF GAS

A shortage of tritium fuel may leave fusion energy with an empty tank

By **Daniel Clery**

In 2020, Canadian Nuclear Laboratories delivered five steel drums, lined with cork to absorb shocks, to the Joint European Torus (JET), a large fusion reactor in the United Kingdom. Inside each drum was a steel cylinder the size of a Coke can, holding a wisp of hydrogen gas—just 10 grams of it, or the weight of a couple sheets of paper.

This wasn't ordinary hydrogen but its rare radioactive isotope tritium, in which two neutrons and a proton cling together in the nucleus. At $30,000 per gram, it's almost as precious as a diamond, but for fusion researchers the price is worth paying. When tritium is combined at high temperatures with its sibling deuterium, the two gases can burn like the Sun. The reaction could provide abundant clean energy—just as soon as fusion scientists figure out how to efficiently spark it.

The interior of the ITER fusion megareactor (artist's concept). It will use up much of the world's tritium.

Last year, the Canadian tritium fueled an experiment at JET showing fusion research is approaching an important threshold: producing more energy than goes into the reactions. By getting to one-third of this breakeven point, JET offered reassurance that ITER, a similar reactor twice the size of JET under construction in France, will bust past breakeven when it begins deuterium and tritium (D-T) burns sometime next decade. "What we found matches predictions," says Fernanda Rimini, JET's plasma operations expert.

But that achievement could be a Pyrrhic victory, fusion scientists are realizing. ITER is expected to consume most of the world's tritium, leaving little for reactors that come after.

ILLUSTRATION: DAVID PARKER/SCIENCE SOURCE

Fusion advocates often boast that the fuel for their reactors will be cheap and plentiful. That is certainly true for deuterium: Roughly one in every 5000 hydrogen atoms in the oceans is deuterium, and it sells for about $13 per gram. But tritium, with a half-life of 12.3 years, exists naturally only in trace amounts in the upper atmosphere, the product of cosmic ray bombardment. Nuclear reactors also produce tiny amounts, but few harvest it.

Most fusion scientists shrug off the problem, arguing that future reactors can breed the tritium they need. The high-energy neutrons released in fusion reactions can split lithium into helium and tritium if the reactor wall is lined with the metal. Despite demand for it in electric car batteries, lithium is relatively plentiful.

But there's a catch: In order to breed tritium you need a working fusion reactor, and there may not be enough tritium to jumpstart the first generation of power plants. The world's only commercial sources are the 19 Canada Deuterium Uranium (CANDU) nuclear reactors, which each produce about 0.5 kilograms a year as a waste product, and half are due to retire this decade. The available tritium stockpile—thought to be about 25 kilograms today—will peak before the end of the decade and begin a steady decline as it is sold off and decays, according to projections in ITER's 2018 research plan.

ITER's first experiments will use hydrogen and deuterium and produce no net energy. But once it begins energy-producing D-T shots, Alberto Loarte, head of ITER's science division, expects the reactor to eat up to 1 kilogram of tritium annually. "It will consume a significant amount of what is available," he says. Fusion scientists wishing to fire up reactors after that may find that ITER already drank their milkshake.

To compound the problem, some believe tritium breeding—which has never been tested in a fusion reactor—may not be up to the task. In a recent simulation, nuclear engineer Mohamed Abdou of the University of California, Los Angeles, and his colleagues found that in a best-case scenario, a power-producing reactor could only produce slightly more tritium than it needs to fuel itself. Tritium leakages or prolonged maintenance shutdowns will eat away at that narrow margin.

Scarce tritium is not the only challenge fusion faces; the field must also learn to deal with fitful operations, turbulent bursts of plasma, and neutron damage (see sidebar, p. 1376). But for Daniel Jassby, a plasma physicist retired from Princeton Plasma Physics Laboratory (PPPL) and a known critic of D-T fusion energy, the tritium issue looms large. It could be fatal for the entire enterprise, he says. "This makes deuterium-tritium fusion reactors impossible."

IF NOT FOR CANDU reactors, D-T fusion would be an unattainable dream. "The luckiest thing to happen for fusion in the world is that CANDU reactors produce tritium as a byproduct," Abdou says. Many nuclear reactors use ordinary water to cool the core and "moderate" the chain reaction, slowing neutrons so they are more likely to trigger fission. CANDU reactors use heavy water, in which deuterium takes the place of hydrogen, because it absorbs fewer neutrons, leaving more for fission. But occasionally, a deuterium nucleus does capture a neutron and is transformed into tritium.

If too much tritium builds up in the heavy water it can be a radiation hazard, so every so often operators send their heavy water to the utility company Ontario Power Generation (OPG) to be "detritiated." OPG filters out the tritium and sells off about 100 grams of it a year, mostly as a medical radioisotope and for glow-in-the-dark watch dials and emergency signage. "It's a really nice waste-to-product story," says Ian Castillo of Canadian Nuclear Laboratories, which acts as OPG's distributor.

Fusion reactors will add significantly to the demand. OPG Vice President James Van Wart expects to be shipping up to 2 kilograms annually beginning in the 2030s, when ITER and other fusion startups plan to begin burning tritium. "Our position is to extract all we can," he says.

But the supply will decline as the CANDUs, many of them 50 years old or more, are retired. Researchers realized more than 20 years ago that fusion's "tritium window" would eventually slam shut, and things have only got worse since then. ITER was originally meant to fire up in the early 2010s and burn D-T that same decade. But ITER's start has been pushed back to 2025 and could slip again because of the pandemic and safety checks demanded by French nuclear regulators. ITER won't burn D-T until 2035 at the earliest, when the tritium supply will have shriveled.

Once ITER finishes work in the 2050s, 5 kilograms or less of tritium will remain, according to the ITER projections. In a worst-case scenario, "it would appear that there is insufficient tritium to satisfy the fusion demand after ITER," concedes Gianfranco Federici, head of fusion technology at the EuroFusion research agency.

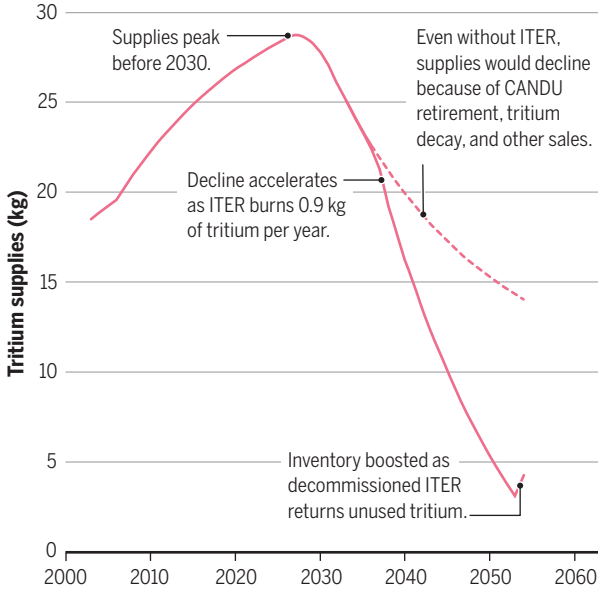
Some private companies are designing smaller fusion reactors that would be cheaper to build and—initially at least—use less tritium. Commonwealth Fusion Systems, a startup in Massachusetts, says it has already secured tritium supplies for its compact prototype and early demonstration reactors, which are expected to need less than 1 kilogram of the isotope during development.

But larger, publicly funded test reactors planned by China, South Korea, and the United States could need several kilograms each. Even more will be needed to start up EuroFusion's planned successor to ITER, a monster of a machine called DEMO. Meant to be a working power plant, it is expected to be up to 50% larger than ITER, supplying 500 megawatts of electricity to the grid.

Fusion reactors generally need a large startup tritium supply because the right conditions for fusion only occur in the hottest part of the plasma of ionized gases. That means very little of the tritium in the doughnut-shaped reactor vessel, or tokamak, gets burned. Researchers expect ITER to burn less than 1% of the injected tritium; the rest will diffuse out to the edge of the tokamak and be swept into a

The dwindling tritium supply

The few kilograms of commercially available tritium come from CANDU plants, a type of nuclear reactor in Canada and South Korea. According to ITER projections, supplies will peak this decade, then begin a steady decline that will accelerate when ITER begins burning tritium.



GRAPHIC: K. FRANKLIN/*SCIENCE*

recycling system, which removes helium and other impurities from the exhaust gas, leaving a mix of D-T. The isotopes are then separated and fed back into the reactor. This can take anywhere from hours to days.

DEMO's designers are working on ways to reduce its startup needs. "We need to have a low tritium [starting] inventory," says Christian Day of the Karlsruhe Institute of Technology, project leader in the design of DEMO's fuel cycle. "If you need 20 kilograms to fill it, that's a problem."

One way to tame the demand is to fire frozen fuel pellets deeper into the reactor's burning zone, where they will burn more efficiently. Another is to cut recycling time to just 20 minutes, by using metal foils as filters to strip out impurities quickly, and also by feeding the hydrogen isotopes straight back into the machine without separating them. It may not be a perfect 50-50 D-T mix, but for a working reactor it will be close enough, Day says.

But Abdou says DEMO's appetite is still likely to be large. He and his colleagues modeled the D-T fuel cycle for power-producing reactors, including DEMO and its successors. They estimated factors, including the efficiency of burning D-T fuel, the time it takes to recycle unburnt fuel, and the fraction of time the reactor will operate. In a paper published in 2021 in *Nuclear Fusion*, the team concludes that DEMO alone will require between 5 kilograms and 14 kilograms of tritium to begin—more than is likely to be available when the reactor is expected to fire up in the 2050s.

EVEN IF THE DEMO team and other post-ITER reactor designers can cut their tritium needs, fusion will have no future if tritium breeding doesn't work. According to Abdou, a commercial fusion plant producing 3 gigawatts of electricity will burn 167 kilograms of tritium per year—the output of hundreds of CANDU reactors.

In May, engineers began to assemble ITER's reactor vessel. The first tritium burns are scheduled for 2035.

The challenge for breeding is that fusion doesn't produce enough neutrons, unlike fission, where the chain reaction releases an exponentially growing number. With fusion, each D-T reaction only produces a single neutron, which can breed a single tritium nucleus. Because breeding systems can't catch all these neutrons, they need help from a neutron multiplier, a material that, when struck by a neutron, gives out two in return. Engineers plan to mix lithium with multiplier materials such as beryllium or lead in blankets that line the walls of the reactors.

ITER will be the first fusion reactor to experiment with breeding blankets. Tests will include liquid blankets (molten mixtures of lithium and lead) as well as solid "pebble beds" (ceramic balls containing lithium mixed with balls of beryllium). Because of cost cuts, ITER's breeder systems will line just 4 square meters of the 600-square-meter reactor interior. Fusion reactors after ITER will need to cover as much of the surface as they possibly can to have any chance of satisfying their tritium needs.

The tritium can be extracted continuously or during scheduled shutdowns, depending on whether the lithium is in liquid or solid form, but the breeding must be relentless. The breeding blankets also have a second job: absorbing gigawatts of power from the neutrons and turning it into heat. Pipes carrying water or pressurized helium through the hot blankets will pick up the heat and produce steam that drives electricity-producing turbines. "All of this inside the environment of a fusion reactor with its ultrahigh vacuum, neutron bombardment, and high magnetic field," says Mario Merola, head of engineering design at ITER. "It's an engineering challenge."

For Abdou and his colleagues, it is more than a challenge—it may well be an impossibility. Their analysis found that with current technology, largely defined by ITER, breeding blankets could, at best, produce 15% more tritium than a reactor consumes. But the study concluded the figure is more likely to be 5%—a worrisomely small margin.

One critical factor the authors identified is reactor downtime, when tritium breeding stops but the isotope continues to decay. Sustainability can only be guaranteed if the reactor runs more than 50% of the time, a virtual impossibility for an experimental reactor like ITER and difficult for prototypes such as DEMO that require downtime for tweaks to optimize performance. If existing tokamaks are any guide, Abdou says, time between failures is likely to be hours or days, and repairs will take months. He says future reactors could struggle to run more than 5% of the time.

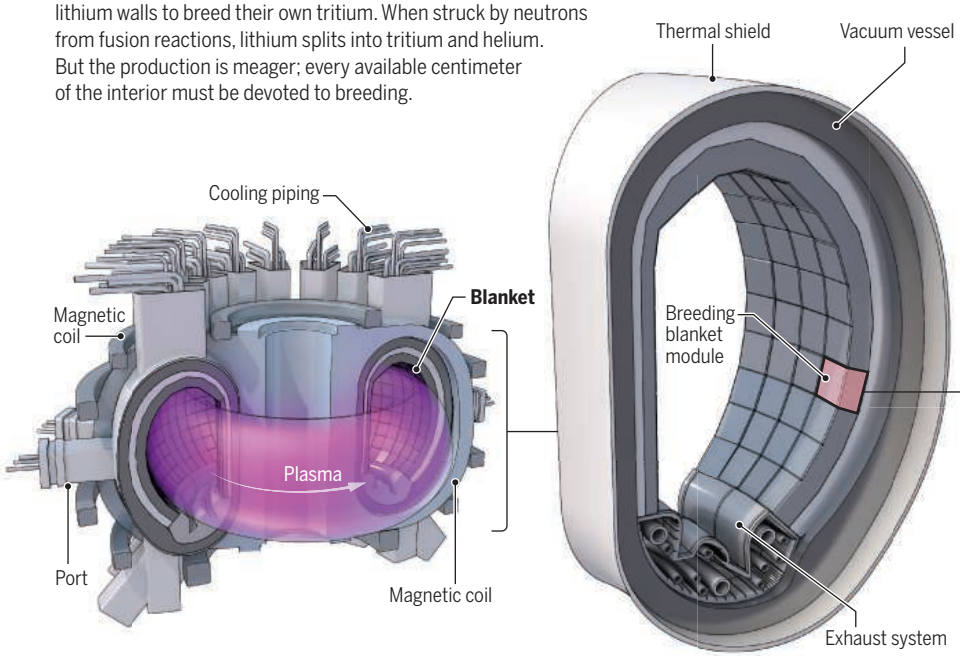
To make breeding sustainable, operators will also need to control tritium leaks. For Jassby, this is the real killer. Tritium is notorious for permeating the metal walls of a reactor and escaping through tiny gaps. Abdou's analysis assumed a loss rate of 0.1%. "I don't think that's realistic," Jassby says. "Think of all the places tritium has to go" as it moves through the complex reactor and reprocessing system. "You can't afford to lose any tritium."

Two private fusion efforts have decided to simply forgo tritium fuel. TAE Technologies, a California startup, plans to use plain hydrogen and boron, whereas Washington state startup Helion will fuse deuterium and helium-3, a rare helium isotope. These reactions require higher temperatures than

PHOTO: © ITER ORGANIZATION

Forging fusion's fuel

The hydrogen isotopes deuterium and tritium are essential fuels for fusion. Whereas deuterium is plentiful, tritium is not. Fusion reactors will need lithium walls to breed their own tritium. When struck by neutrons from fusion reactions, lithium splits into tritium and helium. But the production is meager; every available centimeter of the interior must be devoted to breeding.



Europe's future giant

The European Union is planning DEMO, a fusion reactor bigger than ITER that will demonstrate electricity generation from the 2050s.

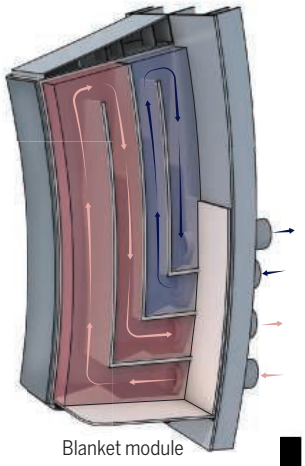
A slice of DEMO

Although designs for DEMO are not set, it will need tritium breeding blankets that could use solid or liquid lithium.

A wet blanket

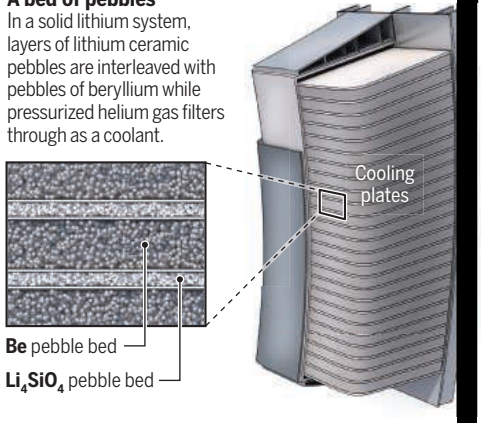
Blankets rely on neutron multipliers such as lead or beryllium (Be), which boost tritium production by spitting out two neutrons for every one absorbed. In a liquid blanket, a molten lead-lithium (PbLi) mix circulates while pressurized water flows past as coolant.

- **PbLi** breeder flow
- **H_2O** coolant flow



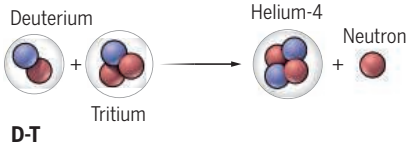
A bed of pebbles

In a solid lithium system, layers of lithium ceramic pebbles are interleaved with pebbles of beryllium while pressurized helium gas filters through as a coolant.

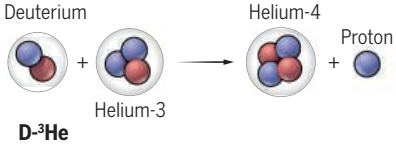


Tritium alternatives

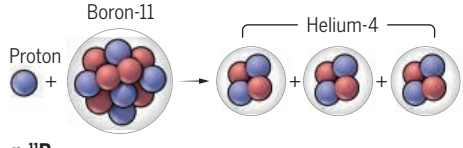
Among fusion fuels, deuterium and tritium burn most easily, but others can work at higher temperatures. The alternatives have the advantage of producing few damaging neutrons or none at all. Also, because their reaction products—protons and helium nuclei—are charged, their energy can directly drive an electric current. In tritium systems, neutrons' heat must be captured to make steam and drive a turbine.



D-T
Deuterium and tritium fuse at 150 million degrees, a temperature within many fusion reactors' reach.



D-^{3}He
Deuterium and helium-3 will fuse at 200 million degrees. But the helium isotope is rare and, like tritium, will need to be bred in the reactor.



p-^{11}B
Hydrogen and boron are both plentiful, but they fuse at 1 billion degrees, which will push reactors into unknown territory.

D-T, but the companies think that's a price worth paying to avoid tritium hassles. "Our company's existence owes itself to the fact that tritium is scarce and a nuisance," says TAE CEO Michl Binderbauer.

The alternative fusion reactions have the added appeal of producing fewer or even no neutrons, which avoids the material damage and radioactivity that the D-T approach threatens. Binderbauer says the absence of neutrons should allow TAE's reactors—which stabilize spinning rings of plasma with particle beams—to last 40 years. The challenge is temperature: Whereas D-T will fuse at 150 million degrees Celsius, hydrogen and boron require 1 billion degrees.

Helion's fuel of deuterium and helium-3 burns at just 200 million degrees, achieved using plasma rings similar to TAE's but compressed with magnetic fields. But helium-3, although stable, is nearly as rare and hard to acquire as tritium. Most commercial sources of it depend on the decay of tritium, typically from military stockpiles. Helion CEO David Kirtley says, however, that by putting extra deuterium in the fuel mix, his team can generate D-D fusion reactions that breed helium-3. "It's a much lower cost system, easier to fuel, easier to operate," he says.

Still, advocates of conventional D-T fusion believe tritium supplies could be expanded by building more fission reactors. Militaries around the world use tritium to boost the yield of nuclear weapons, and have built up their own tritium stockpiles using purpose-built or adapted commercial nuclear reactors.

The U.S. Department of Energy (DOE), for example, relies on commercial reactors—Watts Bar Units 1 and 2, operated by the Tennessee Valley Authority—in which lithium control rods have replaced some of the boron ones. The rods are occasionally removed and processed to extract tritium. DOE supplied PPPL with tritium in the 1980s and '90s when the lab had a D-T burning reactor. But Federici doesn't think the agency, or militaries around the world, will get into the business of selling the isotope. "Defense stockpiles of tritium

GRAPHIC: C. BICKEL/*SCIENCE*

Breakdowns could plague fusion power plants

For decades, achieving controlled fusion was a physics challenge. But now, as the ITER megaproject gears up to demonstrate fusion's potential as an energy source—and startup companies race to beat it—the practical roadblocks to fusion power plants are coming into focus. One is a looming shortage of tritium fuel (see main story, p. 1372). Others could prevent reactors from ever running reliably—a necessity if fusion is to provide a constant "baseload" to complement intermittent solar and wind power.

Some of fusion's fitfulness is innate to the design of doughnut-shaped tokamak reactors. The magnetic field that confines the ultrahot, energy-producing plasma is generated in part by the charged particles themselves, as they flow around the vessel. That plasma current in turn is induced by pulses of electrical current in a coil of wire in the doughnut's hole, each lasting a few minutes at most. In between pulses the magnetic field ebbs, interrupting tokamak operations—and power delivery. The repetitive starts and stops of the reactor's powerful magnetic fields also generate mechanical stresses that could eventually tear the machine apart.

In theory, the beams of particles and microwaves used to heat the plasma can also drive the plasma current. So can a quirk of plasma physics called the bootstrap effect. Near the edge of the plasma, a sharp pressure gradient causes the particles to spiral in such a way that they interfere with each other and push themselves—by their own bootstraps—around the ring.

Using a combination of beams and bootstrap, researchers at ITER think they can get hourlong runs. But the bootstrap effect works best at high pressures and can push the plasma out of control, potentially damaging the reactor, says Alberto Loarte, head of ITER's science division.

Such outbursts of turbulent plasma are another headache for reactor operators, because they can scour metal off the vessel's inner wall, not only threatening its integrity, but also poisoning the plasma. At the Joint European Torus (JET), a U.K.-based tokamak with a reactor wall made of beryllium and tungsten, an automated protection system injects gas into the plasma to staunch the bursts—but not always successfully. "You get drops of beryllium everywhere," says Fernanda Rimini, JET's plasma operations expert.

A turbulent outburst of plasma was caught on camera at the Mega Ampere Spherical Tokamak, a small fusion reactor in the United Kingdom.

ITER operators hope to quell the disruptions by firing frozen deuterium pellets into the plasma and applying an additional magnetic field. Both measures should make the edge of the plasma slightly leaky, so breakouts are small and manageable rather than big and damaging.

The flood of high-energy neutrons produced by fusion reactions pose another threat. The neutrons are a "double-edged sword," says materials scientist Andy London of the UK Atomic Energy Authority. On the one hand, they dump heat in the reactor wall that ultimately generates electricity, and they can bombard lithium to breed tritium fuel. But they can also penetrate the reactor walls and lodge in surrounding steel structures, knocking atoms out of position and weakening the material. Nuclei in the structures sometimes absorb the neutrons, creating radioactive isotopes that do further damage. For example, neutron bombardment can turn the nickel in many steel alloys into a form that gives off helium, causing the steel to swell perceptibly. "The metal turns into a sponge," London says.

Finding tougher materials is a challenge, London says, because "we don't have the luxury of a fusion reactor we can test materials in." A planned European accelerator facility in Spain, dubbed IFMIF-DONES, is supposed to test fusion materials with the world's most intense neutron beam. But construction has not started.

Fixing damaged or weakened reactor components will be slow. Because of the hostile radioactive environment, repairs will rely on robots or remote handling arms that can navigate the narrow access ports of a tokamak. Mohamed Abdou, a nuclear engineer at the University of California, Los Angeles, believes future reactors may operate less than 5% of the time.

Compare this, he says, with today's fission reactors. They can keep running even when individual fuel rods fail. Cranes can swap out fuel rods in just a couple of days. Availability can be as high as 90%. Achieving something similar for fusion will be "very challenging," Abdou says. —*D.C.*

PHOTO: UK ATOMIC ENERGY AUTHORITY

are unlikely ever to be shared," he says.

Perhaps the world could see a renaissance of the CANDU technology. South Korea has four CANDU reactors and a plant for extracting tritium but does not sell it commercially. Romania has two and is working on a tritium facility. China has a couple of CANDUs and India has built a handful of CANDU derivatives. Their tritium production could be turbocharged by adding lithium rods to their cores or doping the heavy water moderator with lithium. But a 2018 paper in *Nuclear Fusion* by Michael Kovari of the Culham Centre for Fusion Energy and colleagues argues such modifications would likely face regulatory barriers because they could compromise reactor safety and because of the dangers of tritium itself.

Some say fusion reactors could create their own startup tritium by running on deuterium alone. But D-D reactions are wildly inefficient at tokamak temperatures and instead of producing energy would consume huge amounts of electricity. According to Kovari's study, D-D tritium breeding might cost $2 billion per kilogram produced. All such solutions "pose significant economic and regulatory difficulties," Kovari says.

Throughout the decades of fusion research, plasma physicists have been single-minded about reaching the breakeven point and producing excess energy. They viewed other issues, such as acquiring enough tritium, just "trivial" engineering, Jassby says. But as reactors approach breakeven, nuclear engineers like Abdou say it's time to start to worry about engineering details that are far from trivial. "Leaving [them] until later would be hugely mistaken." ■

PHOTO: INGO ARNDT/MINDEN PICTURES

OCEANOGRAPHY

Phytoplankton response to a warming ocean

The nutritional value of marine algae will decrease in a warmer world

By **Julio Sepúlveda**[1,2] *and* **Sebastian I. Cantarero**[1]

Phytoplankton, a diverse group of small photosynthetic algae inhabiting the sunlit region near the ocean surface, form the base of marine trophic webs (*1*). Whereas phytoplankton have evolved in tandem with the climate system for hundreds of millions of years (*2*), cumulative greenhouse gas emissions are causing rising ocean temperature, acidification, and oxygen loss at increasing rates (*3*). How phytoplankton will respond and adapt to these multistressors in the future (*4*), and how this will in turn threaten marine trophic webs and food supply for humans, remain important questions in oceanography. On page 1487 of this issue, Holm *et al.* (*5*) show that a physiological adaptation of phytoplankton to a warming world will lead to a reduction in their nutritional value, with negative consequences for marine ecosystems.

Microorganisms have a skillful way of coping with the physiological stress imposed by changing environmental conditions. This mechanism, known as lipid remodeling, involves chemical modification of the lipids responsible for controlling the structure, permeability, and fluidity of the cell membrane of phytoplankton, and thus their interaction with the environment (*6*). One such adaptation mechanism, known as homeoviscous adaptation, alters the number of bonds between carbon atoms in lipids in response to varying temperatures to regulate the fluidity and stability of the cell membrane (*7*). To take advantage of this mechanism, phytoplankton may produce more saturated fats (with a single bond) and fewer unsaturated fats (with one or more double bonds) to maintain the structural integrity of their cell membranes at a higher temperature. For instance, the solid state of butter at room temperature, as opposed to vegetable oil, is

[1]Department of Geological Sciences and Institute of Arctic and Alpine Research (INSTAAR), University of Colorado Boulder, Boulder, CO 80309, USA. [2]Millennium Institute of Oceanography (IMO), University of Concepción, Concepción, Chile. Email: jsepulveda@colorado.edu; sebastian.cantarero@colorado.edu

Phytoplankton (shown here as Diatoms *Rhizosodenia stolterfothii* and *Guinarolia flaccida*) form the base of the marine food web. They will produce more saturated fats at the expense of unsaturated fats in a warmer ocean, which will negatively affect the marine ecosystem.

explained by the larger proportion of saturated fats in butter.

Despite being a well-understood process, few studies have explored the consistency of this specific adaptation across multiple and widely distributed taxonomic groups (*8*) or provided detailed characterizations of the chemically complex lipid classes in the ocean. A notable challenge to investigate homeoviscous adaptation at a large scale is the intensive laboratory procedures and computations required to measure the vast and complex diversity of lipids in the natural environment. Thankfully, with recent advances in environmental lipidomics (*9, 10*), which is a rapidly developing field that studies the lipid pool of an ecosystem, there are now more resources and tools to help researchers overcome some of these challenges.

Using state-of-the-art analytical and computational techniques that allow for the characterization of thousands of molecular structures, Holm *et al.* provide a detailed, global-scale study of lipid remodeling in marine phytoplankton across a wide range of oceanographic settings and temperatures. The authors not only confirm the relationship between temperature and the degree of unsaturation in fatty acids, but also demonstrate how ongoing and projected global warming (*11*) will change the marine phytoplankton lipidome. In particular, a higher proportion of saturated fatty acids is expected in a warmer ocean, which will reduce the availability of essential unsaturated fatty acids to the food web. Because animals cannot produce their own essential unsaturated fatty acids, they must obtain them through their diet. A reduction in the nutritional value of phytoplankton could have negative consequences for fish immune and reproductive systems (*12*), and subsequently affect economically critical fisheries and food supplies for a large fraction of the world's population. Additionally, the vast lipidomic dataset made publicly available by Holm *et al.* represents a wealth of chemical and biological information for oceanographers to study microbial responses through lipid remodeling not only to temperature, but also to multi-environmental stressors—for example, nutrient content, oxygen availability, and acidity—in different regions of the world's ocean.

The development of sophisticated analytical tools in environmental genomics has added an invaluable perspective on biological interactions in the global ocean (*13*), yet a more recent focus has been on the exchange of organic molecules between microorganisms that drive ocean chemistry and productivity (*14*). The study of Holm *et al.* marks the beginnings of a new era in lipidomic research not only in the ocean, but also in other ecosystems. The prospects are exciting, particularly if lipidomics is paired with other fields such as the ecosystem-level study of the complete set of genetic information (metagenomics), actively expressed genes (metatranscriptomics), and metabolic byproducts (metabolomics). Such meta-omics studies will help refine the connection between lipids and specific biological producers and metabolic processes, as well as define the environmental drivers that promote lipid biosynthesis, which will, in turn, improve understanding of their role in cell physiology. Furthermore, lipid remodeling in global-scale observational studies could be further validated if the techniques used by Holm *et al.* are applied in large-scale, community-level experimental studies such as mesocosms (*15*).

To elevate the field of environmental lipidomics to the next level, chemical and biological oceanographers must consolidate analytical and computational procedures and expand and curate shared databases in mass spectrometry (*10*), just as microbiologists have done with genomic techniques in recent decades. Such efforts will facilitate the identification, quantification, and spatial and temporal distribution of previously undescribed molecules to reveal yet unknown biological sources and responses to global-scale environmental perturbations. ■

REFERENCES AND NOTES

1. E. Litchman *et al.*, *J. Ecol.* **103**, 1384 (2015).
2. S. Sharoni, I. Halevy, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2113263118 (2022).
3. P. A. Arias *et al.*, in Climate Change 2021: The Physical Science Basis. Technical Summary. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, V. Masson-Delmotte et al., Eds. (Cambridge Univ. Press, 2021), pp. 33–144.
4. P. W. Boyd, S. T. Lennartz, D. M. Glover, S. C. Doney, *Nat. Clim. Chang.* **5**, 71 (2015).
5. H. C. Holm *et al.*, *Science* **376**, 1487 (2022).
6. A. L. Santos, G. Preta, *Cell. Mol. Life Sci.* **75**, 1909 (2018).
7. R. Ernst, C. S. Ejsing, B. Antonny, *J. Mol. Biol.* **428** (24 Pt A), 4776 (2016).
8. S. M. Hixson, M. T. Arts, *Glob. Change Biol.* **22**, 2744 (2016).
9. J. R. Collins, B. R. Edwards, H. F. Fredricks, B. A. S. Van Mooy, *Anal. Chem.* **88**, 7154 (2016).
10. A. D. Steen *et al.*, *Front. Mar. Sci.* **7**, 718 (2020).
11. L. Kwiatkowski *et al.*, *Biogeosciences* **17**, 3439 (2020).
12. M. T. Arts, M. T. Brett, M. J. Kainz, Eds., in *Lipids in Aquatic Ecosystems* (Springer, 2009). pp. 237–255.
13. E. F. DeLong, *Nature* **459**, 200 (2009).
14. M. A. Moran, *Science* **350**, aac8455 (2015).
15. T. B. Meador *et al.*, *Front. Mar. Sci.* **4**, 113 (2017).

ACKNOWLEDGMENTS

The authors acknowledge support from NSF CAREER award 2047057.

10.1126/science.abo5235

MICROBIOLOGY

A bacterium that is not a microbe

A new discovery challenges the prevailing view of the boundaries of bacterial cell size

By **Petra Anne Levin**

Between ~0.4 and 3 μm^3 in volume, the average bacterium is visible only with a powerful compound microscope. According to the textbook definition, "a microbe is a microscopic organism" [(*1*), p. 6]. Bacteria are thus prototypically microbial. Except when they are not. On page 1453 of this issue, Volland *et al.* (*2*) report the identification of a bacterium that is—astonishingly—visible to the naked eye. At almost 1 cm in length, the Gram-negative proteobacterium *Thiomargarita magnifica* lives attached to sunken leaves in the sulfurous waters of mangrove forests in the archipelago of Guadeloupe. This discovery adds to the group of large sulfur bacteria (LSB) and helps to solve the puzzle of what factors limit cell size.

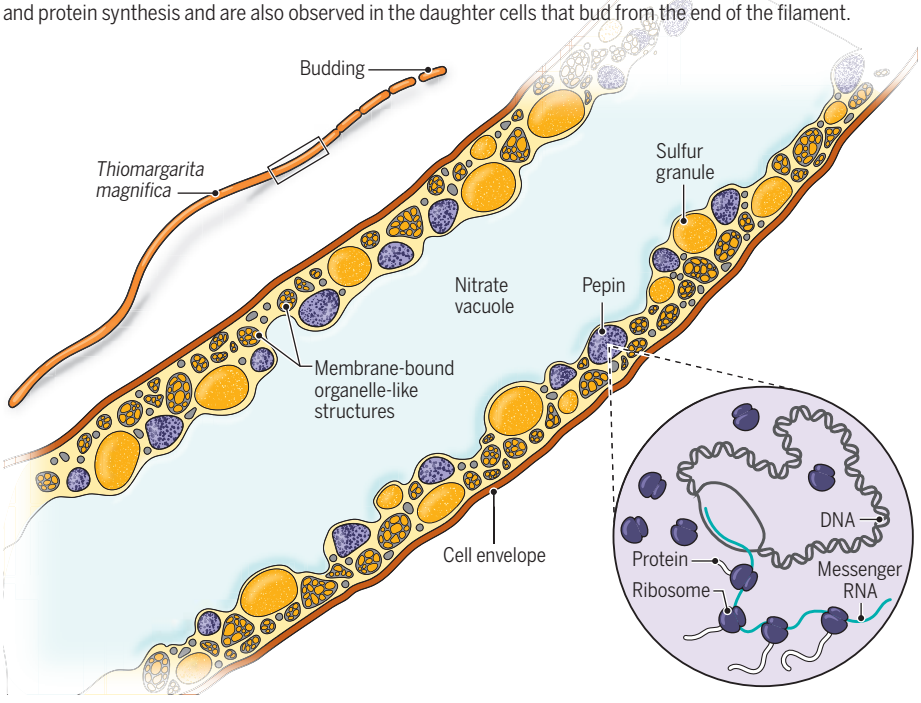
When cells are small, sufficient volume is needed to contain genetic material and carry out basic biosynthesis for growth and proliferation (*3*). For large cells, the diffusion of molecules is likely a primary challenge. Small molecules take an hour to move 1 mm by diffusion, making acquisition of nutrients, dissemination of signals, and disposal of waste products exceedingly difficult at large cytoplasmic volumes (*4*). The size problem is not one without a solution. Eukaryotic cells range widely in size, with volumes between ~1 and 1000 μm^3. For large eukaryotic cells, the problem of diffusion is solved in multiple ways: specialized organelles compartmentalize essential functions, including transcription, adenosine triphosphate (ATP) synthesis, and secretion; and transport systems efficiently move nutrients, cellular building blocks, and waste products to the correct location.

Bacteria have also solved the size problem (*5*). Living in a sulfur-rich environment,

Department of Biology, Washington University in St. Louis, St. Louis, MO, USA. Email: plevin@wustl.edu

Thiomargarita magnifica

The large vacuole of *Thiomargarita magnifica* is likely filled with nitrate, and the thin layer of cytoplasm contains sulfur granules, which are involved in metabolism. Adenosine triphosphate (ATP) synthases localize to the intracytoplasmic membrane network, functionally increasing the surface area available for generating ATP. Also embedded in the cytoplasm are "pepins." These membrane-bound vesicles are the site of RNA and protein synthesis and are also observed in the daughter cells that bud from the end of the filament.



T. magnifica oxidizes hydrogen sulfide and reduces nitrate; the nitrate is likely to be stored in a large vacuole that accounts for ~75% of total cell volume. In this regard, *T. magnifica* is not unlike its more diminutive LSB relatives, particularly *Thiomargarita namibiensis* (*6*). As the previous record holder for large bacteria (average diameter ~750 μm), 98% of the volume of a *T. namibiensis* cell is occupied by a large nitrate-containing vacuole.

How do *T. magnifica* and other giant bacteria solve the diffusion problem? The presence of the nitrate vacuole in LSBs is part of the answer. Pressed up against the cell envelope by the vacuole, the cytoplasmic layer in *T. magnifica* is only ~2 to 3 μm in depth, facilitating the diffusion and transport of nutrients and other molecules into the cell and the export of waste products out. In addition to sulfur granules that are necessary for respiration, the cytoplasm contains a complex network of membranes. The F_1F_0 ATP synthase, which is responsible for generating ATP through proton motive force, localizes to this network, suggesting another mechanism by which *T. magnifica* overcomes challenges associated with a lower surface area–to-volume ratio.

Polyploidy is also important for overcoming diffusion limits (*5*). Distributed genomes allow for localized transcription and translation, eliminating the need to transport proteins across vast distances to fulfill basic biosynthetic needs. Large, metabolically active cells are typically polyploid. *Drosophila melanogaster* salivary glands contain hundreds of chromosomal copies, as do some specialized plant cells. *Epulopiscium fishelsoni*, a member of the Gram-positive firmicutes whose length can exceed 600 μm, contains hundreds of thousands of genome copies localized to the cell periphery (*7*, *8*). *T. magnifica* is estimated to contain ~40,000 genome copies.

Notably, *T. magnifica* takes polyploidy a step further than other LSBs, neatly packaging its genome, or perhaps genomes, into membrane-bound vesicles distributed throughout the cytoplasm. These compartments (called pepins) contain not only DNA but also ribosomes (see the figure). Together with data indicating that they are a primary site of protein synthesis, their structure suggests that pepins function almost as autonomous organisms within the larger cell, conducting simultaneous transcription and translation, as is common in bacteria. Reaffirming the importance of the pepins to the *T. magnifica* life cycle, Volland *et al.* also observed them in the daughter cells that bud from the ends of long filaments, in what the authors speculate is a form of reproduction. A slow metabolism appears to be an additional adaptation to the large size of *T. magnifica*. The authors predict that *T. magnifica* will require up to 2 weeks to produce daughter cells. Labeling indicates the presence of "hot spots" for protein synthesis near sites of constriction at cell poles, which suggests a mechanism to direct activity to sites of growth and differentiation.

Despite their enormous size, the surface of *T. magnifica* is notably pristine. Biofilms (bacterial communities) form on any surface large enough to support a microbe. Although conservatively >625,000 *Escherichia coli* could fit on the surface of a single *T. magnifica* cell, scanning electron microscopy indicates that their surface is devoid of epibiotic bacteria. Biosynthetic gene clusters, many of which encode nonribosomal peptide synthases and polyketide synthases, account for ~25% of the large 12-Mb genome of *T. magnifica*. This distribution is reminiscent of biosynthetic gene clusters in the genome of the actinomycetes, which are a major source of antibiotics and antibiotic precursors. These biosynthetic gene clusters—and particularly their products—provide a likely explanation for the lack of surface-associated organisms.

How such a large bacterium stimulates biosynthesis at specific subcellular regions to coordinate growth and development is a mystery. The mechanisms that control the import of metabolites and export of waste products, as well as responses to environmental stress, remain opaque. Illuminating these and other aspects of *T. magnifica* biology will require developing methods to grow it in culture, something that has been achieved for only a small fraction of bacterial species because of fastidious and frequently elusive growth requirements. Why these organisms need to be so large is another equally intriguing, if challenging, issue. More philosophical in nature is the question of whether *T. magnifica* represents the upper limit of bacterial cell size. It seems unlikely, and, as the study of Volland *et al.* illustrates, bacteria are endlessly adaptable and always surprising—and should never be underestimated. ■

REFERENCES AND NOTES

1. J. Slonczewski, J. W. Foster, *Microbiology: An Evolving Science* (W. W. Norton & Company, ed. 4, 2016).
2. J.-M. Volland *et al.*, *Science* **376**, 1453 (2022).
3. A. L. Koch, *Annu. Rev. Microbiol.* **50**, 317 (1996).
4. H. N. Schulz, B. B. Jorgensen, *Annu. Rev. Microbiol.* **55**, 105 (2001).
5. P. A. Levin, E. R. Angert, *Cold Spring Harb. Perspect. Biol.* **7**, a019216 (2015).
6. H. N. Schulz *et al.*, *Science* **284**, 493 (1999).
7. E. R. Angert *et al.*, *Nature* **362**, 239 (1993).
8. J. E. Mendell *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6730 (2008).

ACKNOWLEDGMENTS

Work in the Levin lab is supported by a National Institutes of Health grant to P.A.L. (R35 GM127331).

10.1126/science.adc9387

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

The many faces of anthropogenic subsidence

Shallow and deep human subsurface activities contribute to the total subsidence

By **Thibault Candela** and **Kay Koster**

Human-induced subsidence clearly causes social problems, is costly, and poses flooding and damage threats (*1*). Less evident is how much each spatially and temporally superimposed human subsurface activity contributes to total subsidence. Subsidence is often caused by an accumulation of multiple shallow and deep subsurface activities, such as groundwater withdrawal and hydrocarbon extraction (*2*). This complex challenge is well known but has not been properly addressed. A turning point has now been reached at which advanced monitoring strategies and physics-based models can be used in concert to clarify the subsidence contribution of each human intervention. Identifying these causal relationships between human interventions and subsidence helps to establish overarching responsibilities and to better develop science-based site-specific prevention and mitigation strategies.

Multiple human interventions spanning a wide range of depth can cause subsidence. The main documented superimposed subsurface activities causing subsidence are shallow groundwater withdrawal and deep hydrocarbon extraction. The Netherlands is a typical coastal area where research efforts have considered shallow and deep human interventions that cause subsidence as isolated boxes. Studies of the subsidence caused by deep gas production at the Groningen field (*3*) disregard the potential contribution of groundwater management. In parallel, multiple studies have documented subsidence that is induced by groundwater management (*4*), disregarding the potential contribution of hydrocarbon extraction. A similar picture emerges in other coastal areas, such as the Mississippi Delta, where most of the attention has been focused on the effects of groundwater withdrawal (*5*) and little attention has been given to the abundantly present nearby hydrocarbon fields (*6*). However, the superimposition of human-controlled sources of subsidence is not only occurring in the Netherlands and the Mississippi Delta. For some cases, the combination of groundwater withdrawal at relatively shallow levels and deep gas extraction contribute to the total subsidence, such as the Northern Adriatic coastland, including Venice lagoon, the Niger delta, and Indonesia (*2*). For other cases, such as the Mekong delta (*7*), local subsidence is caused by groundwater management at multiple depths. For example, the extraction of groundwater from shallow-depth and mid-depth aquifers for domestic and industrial uses, respectively.

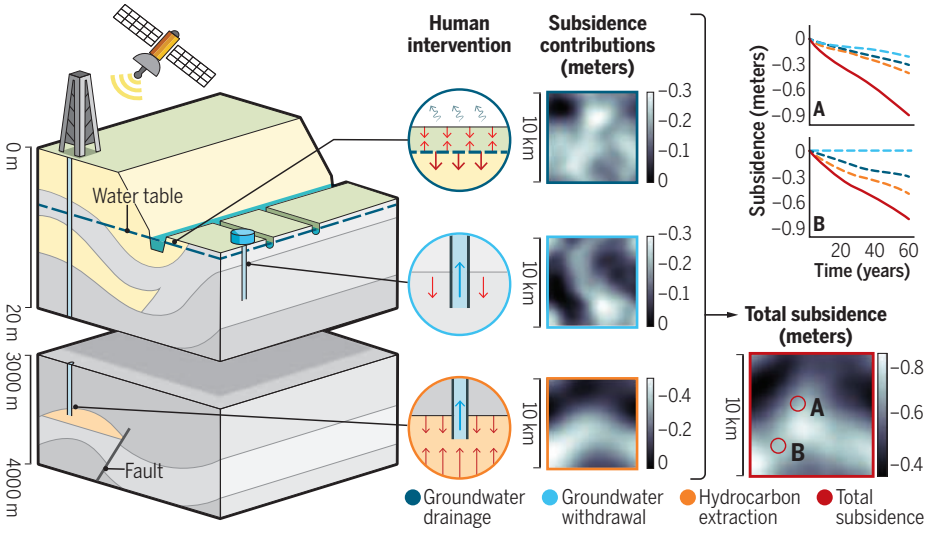
For deep hydrocarbon and shallower groundwater withdrawal, lowering of ground surface elevation occurs in cases when producing fluid from the subsurface decreases pressure in the pore space of the rock that leads to subsurface compaction (see the figure) (*2*). Additional human intervention at even very superficial depth can lead to subsidence. For example, when draining very superficial groundwater, organic rich soils (peat) can be exposed to the air and trigger microorganisms to deteriorate the organic carbon, ultimately leading to severe greenhouse gas emissions and subsidence (*4*). In the Netherlands, this process has already been identified and extensively studied. In other parts of the world, such as the Mississippi Delta, the importance of the very shallow soil layer only recently has been recognized in the overall total subsidence budget (*8*). Voosen (*8*) recently reported that many river deltas are sinking faster than previously believed because of the failure to account for the compaction of these very shallow soil layers.

Monitoring and ultimately deconvolving the different contributions of subsidence is challenging. Using conventional geodetic techniques, the measurement of the compaction of very shallow soil layers is notoriously difficult because of the scarcity of fixed benchmarks installed in such soft soil. Hoyt *et al*. (*9*) overcame this challenge and demonstrated that satellite-based interferometric synthetic aperture radar (InSAR) measurements can monitor compaction of very shallow soil layers over the Southeast Asian peatlands. Chaussard *et al*. (*10*) demonstrated that by further post-processing the InSAR measurements with statistical tools (independent component analysis), separation of the multiple components of the total subsidence signal is possible, as well as the linking of each component to a specific subsurface activity.

Instead of relying only on InSAR measurements, multiple observation techniques are being used in concert at the Mississippi Delta and the Venice Lagoon (*11*). InSAR measurements are combined with a fit-for-

Subsidence through multiple human interventions

Modern observation can map surface subsidence over an area of concern (left column) and help disentangle the relative contributions from multiple causes (middle column). Disregarding any potential human intervention will risk overestimating the role of others at any point on the surface (for example, points A and B in a 10 by 10 km surface) as total subsidence (red line) accumulates over time.



Geological Survey of the Netherlands, Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek (TNO), Utrecht, Netherlands.
Email: thibault.candela@tno.nl

GRAPHIC: NIKKI TRABUCHO/TNO, GEOLOGICAL SURVEY OF THE NETHERLANDS, UTRECHT, THE NETHERLANDS

purpose surface elevation table apparatus and with at-depth along-wells optical fiber strainmeters. Following this integrated approach, the along-wells deeper measurements exclusively record the subsidence contribution of deep human interventions. By contrast, the InSAR measurements of the changes in elevation of the ground surface record both deep and shallow contributions. In the Netherlands, combining diverse geodetic techniques (absolute gravity, Global Navigation Satellite System, and InSAR) has allowed separation of the measuring points into two families (*7*). One monitors deep human interventions such as hydrocarbon extraction, and the other monitors shallow human interventions such as groundwater level changes.

The subsidence footprint of each human intervention can now be measured with tailored monitoring technologies. This development is a major step forward, which helps avoid the tunnel vision that can lead to erroneously attributing the total subsidence to one single human subsurface activity. Because these tailored monitoring technologies are in place, unbiased assessment of the subsidence rate caused by each superimposed human subsurface activity can be deployed.

Connecting the underlying physical causes of subsidence to the observation-based disentanglement allows a full picture of the origin of subsidence. Subsurface modeling is required because of the limited predictive power of observations but allows for establishing causal relationships. After demonstrating that the models can explain the observations, these validated models can be used to test targeted prevention and mitigation strategies for each human intervention.

Data assimilation techniques designed to calibrate physical models with observations, identical to those used for weather forecasting, have recently been successfully deployed for human-induced subsidence forecasting (*4*, *12*). However, these studies have only focused on a single subsurface activity at a time. The natural evolution would be to use the same data assimilation techniques to jointly calibrate, in one go, models of subsidence induced by shallow and deep human interventions. This calibration step will help to discriminate deep-seated causes and shallow-seated causes from the total subsidence signal. By tightly connecting observations and models, physical explanation will support the observation-based disentanglement.

The negative effects of human-induced subsidence are known, especially in terms of financial consequences. City officials of Amsterdam, the Netherlands, for example, estimate their municipality to suffer more than €60 million per year worth of subsidence-related building damage and infrastructural maintenance costs (*13*). However, the sense of emergency is difficult to perceive because subsidence is a slow and gradual motion. This results in a passive decision-making process that may be exacerbated by attaching each superimposed subsurface activity to a specific actor. This makes it unclear to what extent an actor may be responsible for the total subsidence. Disentangling the total subsidence into the relative contribution of each human intervention could allow for a more active mode for planning mitigation procedures rather than the more passive method of focusing on adaptation strategies.

Underlying the need for disentangling subsidence source is the recognition that human interventions in the subsurface will be ongoing for the foreseeable future. Large-scale groundwater withdrawal for domestic, industrial, and agricultural use will continue and possibly expand with changing weather patterns and increasing population. Compaction of hydrocarbon reservoirs continues for decades even after production stops, so any decreases in hydrocarbon extraction will not immediately stop the related subsidence. The potential also exists for large-scale human interventions to ramp up for long-term underground CO_2 storage and hydrogen, along with geothermal energy extraction. All of these can lead to subsidence and will need to be individually accounted for by establishing causal relationships (*14*). The issues are especially crucial because coastal cities are already exposed to the threats of global sea-level rise and collectively affected by subsidence-induced groundwater management (*15*). New sources of human-induced subsidence put these places in an even more vulnerable position. ■

REFERENCES AND NOTES

1. G. Herrera-García *et al.*, *Science* **371**, 34 (2021).
2. M. Shirzaei *et al.*, *Nat. Rev. Earth Environ.* **2**, 40 (2021).
3. J. Smith *et al.*, *J. Geophys. Res. Solid Earth* **124**, 6165 (2019).
4. P. A. Fokker, J. L. Gunnink, K. Koster, G. Lange, *J. Geophys. Res. Earth Surf.* **124**, 1099 (2019).
5. C. E. Jones *et al.*, *J. Geophys. Res. Solid Earth* **121**, 3867 (2016).
6. J. W. Day, H. Clark, C. Chang, R. Hunter, C. Norman, *Water* **12**, 1492 (2020).
7. X. Yan *et al.*, *Acta Geol. Sin. Engl. Ed.* **94**, 162 (2020).
8. P. Voosen, *Science* **363**, 441 (2019).
9. A. M. Hoyt, E. Chaussard, S. S. Seppalainen, C. F. Harvey, *Nat. Geosci.* **13**, 435 (2020).
10. E. Chaussard *et al.*, *J. Geophys. Res. Solid Earth* **126**, e2020JB020648 (2021)
11. C. Zoccarato, C. Da Lio, *Nat. Commun. Earth Environ.* **2**, 77 (2021).
12. L. Gazzola *et al.*, *Computat. Geosci.* **25**, 1731 (2021).
13. T. Pelsma *et al.*, *Proc. IHAS* **382**, 669 (2020)
14. K. Im, J.-P. Avouac, E. R. Heimisson, D. Elsworth, *Nature* **595**, 70 (2021).
15. P. C. Wu *et al.*, *Geophys. Res. Lett.* **49**, 7 (2022).

10.1126/science.abn3676

CHEMISTRY

Nitriles for the production of various amines

Nitriles functionalize amines and ammonia under catalytic hydrogenation conditions

By **Yasunari Monguchi**

Amines are commercially important chemicals that are widely used in pharmaceuticals and agrochemicals, with an ever-growing demand for their industrial production (*1*, *2*). Catalytic hydrogenation, which uses hydrogen gas as a reducing reagent, is a well-established industrial process for producing various chemicals. There are many benefits of this method. For instance, the hydrogen gas can be removed by simply opening the reaction vessel; there is no need for removing unreacted reagents and reagent-derived wastes. However, it has been difficult to apply hydrogenation to produce specific amines because the reaction lacks control over selectivity. On page 1433 of this issue, Chandrashekhar *et al.* (*3*) demonstrate a general method to produce various amines using a catalytic hydrogenation condition in a highly selective manner.

Amines are nitrogen-containing compounds that are derivatives of ammonia (NH_3), with one or more hydrogen atoms replaced with carbon substituents from either aliphatic compounds, that is, from the alkyl group, or aromatic compounds, that is, from the aryl group. Amines have traditionally been produced by substituting the hydrogen atoms in ammonia with substituents through chemical reactions. This process is known as alkylation when the substituent is from the alkyl group. However, it is practically impossible to control the number of introduced substituents for each ammonia molecule using this method, resulting in the generation of all substituted amines, that is, ammonia molecules with one (primary), two (secondary), or all three (tertiary) of its hydrogen

Laboratory of Organic Chemistry, Daiichi University of Pharmacy, 22-1 Tamagawa-machi, Minami-ku, Fukuoka 815-8511, Japan. Email: monguchi@daiichi-cps.ac.jp

atoms substituted. Another downside of this synthesis method is the toxicity and corrosivity of the alkyl halides that are needed for the process. A different method for amine preparation is the use of carbonyl compounds as alkylating reagents together with hydride-based reducing reagents. However, carbonyl compounds are highly reactive, and such reducing reagents produce a large amount of various inorganic salts after the reaction, thereby requiring complicated and laborious operations. Thus, there is a great demand for an alternative method for the industrial production of amines.

The catalytic hydrogenation of nitriles to the corresponding amines is a hot research area in organic chemistry, especially on the topic of how to exert control over the output ratio of primary, secondary, and tertiary amines (*4*). Recently, several protocols have been developed to accomplish this (*5*–*7*). These methods (*6*, *7*) adequately produce symmetrical secondary and tertiary amines that bear the same substituents on the nitrogen atoms. By contrast, the method established by Chandrashekhar *et al.* can generate unsymmetrical secondary and tertiary amines in a highly selective manner through the functionalization of amines using nitriles as alkyl donors under catalytic hydrogenation conditions.

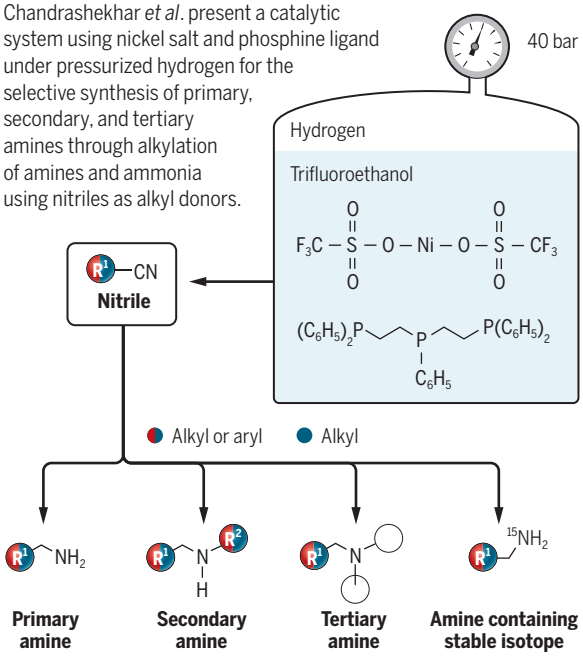
Nitrile is commonly denoted as R^1CN, where R^1 indicates an aliphatic or aromatic substituent that is different from the R^2 aliphatic or aromatic substituent. When nitrile (R^1CN) and primary amine (R^2NH_2) are present with a catalyst and hydrogen gas, the nitrile is first hydrogenated to the corresponding primary imine $R^1CH{=}NH$, which is then transformed to the primary amine $R^1CH_2NH_2$ in a second hydrogenation. If the primary imine reacts with the generated primary amine $R^1CH_2NH_2$ instead of the other amine R^2NH_2, a corresponding symmetrical secondary amine $(R^1CH_2)_2NH$ would form. For the generation of unsymmetrical secondary amine $(R^1CH_2)R^2NH$, the reaction of the coexisting amine R^2NH_2 to the imine $R^1CH{=}NH$ should preferentially occur. As a result, the hydrogenation of the imine should be slow enough that the coexisting amine R^2NH_2 can react with the imine instead of the $R^1CH_2NH_2$.

Although these processes are well understood, reports for the hydrogenative alkylation of amines using nitriles have been limited because of the difficulty of controlling the reaction pathways (*8*–*11*). Chandrashekhar *et al.* use a catalytic system using nickel(II) trifluoromethanesulfonate [$Ni(OTf)_2$, where Tf is CF_3SO_2] as a catalyst and bis(2-diphenylphosphinoethyl) phenylphosphine [$(Ph_2PCH_2CH_2)_2PPh$, where Ph is C_6H_5] as a ligand. The combination worked for the selective generation of unsymmetrical amine through the coupling of a specific nitrile with amine in trifluoroethanol at 100°C under 40 bar of hydrogen (see the figure). According to their experiments, the method can be used to produce a wide diversity of secondary and tertiary amines, including 13 pharmaceuticals, from various amines and nitriles.

Besides secondary and tertiary amines, Chandrashekhar *et al.* demonstrate that their method can be used for the selective synthesis of primary amines by alkylation of ammonia in place of amines. They were able to obtain a variety of primary amines, including fatty amines, in excellent yields. It was also found that the labile functional groups, such as aromatic halides, on nitrile molecules under hydrogenation conditions are untouched during the reaction. Meanwhile, because nitriles have at least two carbon atoms, they cannot be used as a straightforward methyl (CH_3) donor that consists of a single carbon atom for the N-methylation of amines. The authors resolved the problem by using amines that bear methyl substituents preinstalled on nitrogen atoms for the reaction with nitriles as alkyl donors. As a result, an array of *N*-methylamines could be synthesized, although the strong affinity of these amines to nickel ions was suspected to inhibit catalysis.

Amines, obtained from nitriles and amines

Chandrashekhar *et al.* present a catalytic system using nickel salt and phosphine ligand under pressurized hydrogen for the selective synthesis of primary, secondary, and tertiary amines through alkylation of amines and ammonia using nitriles as alkyl donors.



R^1, derived from nitrile; R^2, R^3, and R^4, derived from starting amine.

The method introduced by Chandrashekhar *et al.* can be applied to a so-called late-stage functionalization strategy, that is, the introduction of specific functionalities to complex molecules in the last or in an additional step during synthesis. The strategy is useful, especially for creating derivatives of biologically active compounds in medicine (*12*, *13*). This is because the single chemical modification of known drug molecules could be used to quickly produce new analogs, which are expected to be bioactive as future drug candidates. The authors transformed seven bioactive nitriles to the corresponding primary, secondary, and tertiary amines, which may exhibit biological activities, through the reaction with ammonia or amines. Finally, a larger-scale synthesis of five amines up to 10 g was carried out, pointing to possible industrial-scale applications. Chandrashekhar *et al.* demonstrate that the combined use of nitriles and amines or ammonia in an inexpensive nickel-linear triphos catalytic hydrogenation system can produce a broad range of amines, including pharmaceuticals. Stable isotope–containing amines are also synthesized in a similar system using a reagent derived from stable isotope–containing ammonia and acetic acid, instead of amines or ammonia. A limitation of the method is that less-reactive amines cannot be used as starting amines. In addition, the method can only be applied at plants with high-pressure reaction facilities. These may be possible improvements to tackle in the future. Nevertheless, industrial application of the current catalytic system would be expected because it yields easy access to a variety of desired amines. ■

REFERENCES AND NOTES

1. S. A. Lawrence, *Amines: Synthesis, Properties and Applications* (Cambridge Univ. Press, 2004).
2. O. L. Afanasyev, E. Kuchuk, D. L. Usanov, D. Chusov, *Chem. Rev.* **119**, 11857 (2019).
3. V. G. Chandrashekhar, W. Baumann, M. Beller, R. V. Jagadeesh, *Science* **376**, 1433 (2022).
4. J. A. Garduño, J. J. García, *ACS Catal.* **10**, 8012 (2020).
5. A. Mukherjee, D. Srimani, S. Chakraborty, Y. Ben-David, D. Milstein, *J. Am. Chem. Soc.* **137**, 8888 (2015).
6. K. Rajesh, B. Dudle, O. Blacque, H. Berke, *Adv. Synth. Catal.* **353**, 1479 (2011).
7. Y. Monguchi *et al.*, *J. Org. Chem.* **82**, 10939 (2017).
8. H. Sajiki, T. Ikawa, K. Hirota, *Org. Lett.* **6**, 4977 (2004).
9. T. Ikawa *et al.*, *Org. Biomol. Chem.* **10**, 293 (2012).
10. S. Lu, J. Wang, X. Cao, X. Li, H. Gu, *Chem. Commun.* **50**, 3512 (2014).
11. J. C. Babón, M. A. Esteruelas, A. M. López, E. Oñate, *Organometallics* **39**, 2177 (2020).
12. J. Börgel, T. Ritter, *Chem* **6**, 1877 (2020).
13. K.-H. Huang, J. Ghosh, S. Xu, R. G. Cooks, *ChemPlusChem* **87**, e202100449 (2022).

10.1126/science.abq7782

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

AGING

How ubiquitous is aging in vertebrates?

Two new studies find little evidence of aging in some turtle species

By **Steven N. Austad**[1] *and* **Caleb E. Finch**[2]

The exceptional longevity of chelonians (turtles) has long been appreciated, with reliable although anecdotal reports of individuals surviving for more than 150 years, which raises the question of whether they age at all. On pages 1466 and 1459 of this issue, da Silva *et al.* (*1*) and Reinke *et al.* (*2*), respectively, evaluate survival data from various chelonians and other ectothermic (cold-blooded) tetrapods, including amphibians, snakes, crocodilians, and the tuatara. They find scant evidence of demographic senescence (increased mortality with age) in a number of turtles and several other species in either zoo or wild populations.

The rate of aging, as defined by both studies, is the mortality rate increase with age, calculated as the Gompertz exponential rate constant of annual mortality (*3*). The lack of observed mortality increases with age in many chelonians and amphibians is described as negligible senescence (*4*, *5*), which may be considered perpetual youth. da Silva *et al.* examined the mortality rate changes with age of 52 chelonian species in zoo populations, and aging was not detected in 75% of these species. Moreover, an even higher percentage of both sexes have slower rates of aging than that of humans. These results from zoo animals are similar to conclusions from Reinke *et al.*, who analyzed mark-recapture data in long-term field studies of 77 species of ectothermic tetrapods, including 14 species of chelonians as well as amphibians, snakes, crocodilians, and the tuatara. They also found little evidence of aging in multiple chelonian species as well as in some salamanders and the tuatara. These mortality rate data also fit intuitively with the slow metabolism of chelonians (e.g., Galápagos tortoises have resting heart rates of six beats per minute) as well as the protection from extrinsic dangers offered by their shells (*6*).

The results of da Silva *et al.* and Reinke *et al.* challenge the conclusion of some evolutionary models that senescence (physical aging) is evolutionarily inescapable for any species, with a clear distinction between somatic and germ cell lineages. From both theoretical and practical perspectives, there is considerable distance between the slow aging of chelonians and no aging at all (that is, eternal youth). Could all deaths in these species be the result of infections or accidents that are not related to age? Considerable uncertainty remains about the validity of using data from zoos to test evolutionary theories because these species did not evolve in such protected environments. For three of the species in the zoo study of da Silva *et al.* that showed no aging, field data were also available (*7*–*9*). These data did show aging, emphasizing how wild populations often tell a different story than zoo populations.

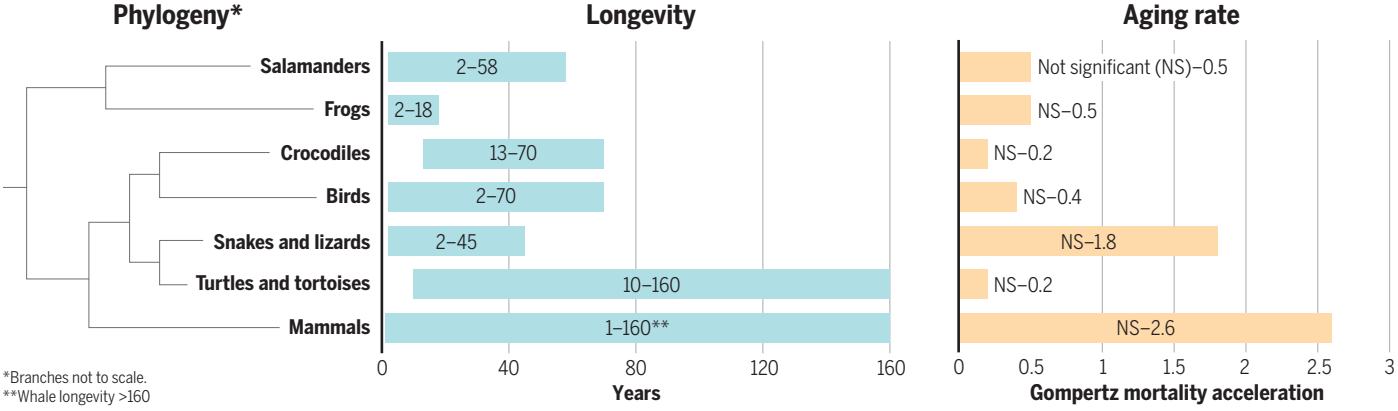
Aging involves more than increasing mortality with age—it also depends on physiological maintenance. For evolutionary theories, mortality is only important to Darwinian fitness because it affects reproduction. Whether reproductive rates are maintained with age is important to ascertaining organismal senescence. Although the studies of da Silva *et al.* and Reinke *et al.* lacked data on age-specific reproductive rates, it is well known that although some chelonians exhibit increased egg production with age, they also have reduced egg hatching (fecundity). Thus, the actual number of live offspring produced may decline even as egg number increases as they grow older (*7*, *10*). Egg hatching data for the putatively nonaging chelonian species are lacking, so it remains possible that some species exhibit neither mortality increase nor fecundity decrease at later ages—that is, they do not age.

Another issue that da Silva *et al.* address is sex differences in survival. In mammals, females often, although not always, are the longer-lived sex. In turtles, it appears to be the reverse. Males in their analysis both live longer and age at a slower rate than females, although this pattern depends on size dif-

[1]University of Alabama at Birmingham, Birmingham, AL, USA. [2]University of Southern California, Los Angeles, CA, USA. Email: austad@uab.edu; cefinch@usc.edu

Longevity of tetrapods

Longevity and aging rate (the Gompertz exponential rate constant of annual mortality) data from da Silva *et al.* and Reinke *et al.* compared with data from mammals and birds reveal that ectothermic and endothermic vertebrates vary greatly in their aging rates. Because the background mortality rate of adults may be high in natural populations, few individuals may reach later ages to show increased mortality. Based on data from (*1, 2, 4–6, 13*).



*Branches not to scale.
**Whale longevity >160

GRAPHIC: K. FRANKLIN/*SCIENCE*

ferences between the sexes. Generally, the smaller sex lives longer. In mice, reduced growth hormone and insulin-like growth factor (IGF) activity reduces size and increases longevity (*11*), so perhaps a similar mechanism is working here.

Reinke *et al.* focus on other hypotheses about variables that affect the rate of aging, comparing the slow turtle senescence with that of other groups. For example, environmental temperature correlated negatively with aging rates of reptiles, as expected from the slowing of metabolism in ectotherms with decreasing environmental temperature. However, this correlation was reversed in amphibians, a finding that deserves some mechanistic follow-up. Among all terrestrial vertebrate groups, turtles had the slowest rate of aging—more than 20-fold slower than that of mammals and twice as slow as that of birds and humans. Notably, the tuatara—a 1- to 2-kg reptile that lives on islands off New Zealand—was found to have a 90% slower aging rate than that of turtles in general and a mean lifespan of 137 years. If confirmed, tuataras would be the longest-lived terrestrial vertebrate.

Overall, these studies suggest that ectothermic tetrapods generally, and turtles in particular, age very slowly, if at all. However, there are some limitations of these data. If basal mortality rates are high—such as what might happen in zoos, where husbandry for a species may not be optimal—life expectancy can be short even with no increase in mortality with age. This is true of much of the data from da Silva *et al.*: Despite no age-related increase in mortality, most turtle life expectancies are just 10 to 20 years, in spite of reliable records of much longer-lived individuals in those species. Indeed, for seven species, mean life expectancy is shorter than the age at first reproduction. The longest-lived of zoo species is the male Aldabra tortoise, with a mean life expectancy of 55.5 years, which is lower than that of humans in high-income countries. Field data on 14 chelonian species from Reinke *et al.* indicate mean life expectancy is much longer: 39 years, measured from age at first reproduction until 95% of the animals had died.

Another limitation, inherent in mortality rate analysis for nearly any species other than humans, is the paucity of individuals at truly advanced ages, where mortality rate could show acceleration. One analysis of mortality rates showed that most laboratory animal studies lacked sufficient numbers of older individuals to assess whether mortality rate was increasing at later ages (*12*). Calculations indicate that the naked mole rat, a mouse-size rodent with very slow aging, would require a population far larger than the current 3300 individuals to detect late-life mortality acceleration (*13*).

It is clear that ectothermic and endothermic vertebrates vary greatly in their aging rates, despite their fundamentally similar cytological and biochemical architectures (see the figure). By investigating the nature of that variation, something new may be learned about aging in humans. If some species truly escape aging, and mechanistic studies may reveal how they do it, human health and longevity could benefit. The naked mole rat is being scrutinized, but no distinct process has been discovered. Even if many of these fascinating species lack significantly increasing mortality with age, some clearly incur infirmities of aging. Diseases of aging, such as cataracts and various cancers, are well known in chelonians. Harriet, a famous Galápagos tortoise that spent her last years in the Australia Zoo, reputedly lived 170 years, but in the end, she died from a heart attack, a well-known malady of human aging. And the oldest living terrestrial animal, Jonathan—an Aldabra tortoise that has allegedly lived on Saint Helena Island since 1882 and therefore may be 160 to 190 years old (birth date was estimated)—is now blind, has lost his olfactory sense, and must be fed by hand. Demography does not always perfectly reflect physiological decline, particularly in highly protected environments like zoos. So whatever the demography of that species suggests, these ancient tortoises clearly have aged. Thus, the observed longevity may not always correlate with the aging rate. ■

"It is clear that ectothermic and endothermic vertebrates vary greatly in their aging rates..."

REFERENCES AND NOTES

1. R. da Silva, D. A. Conde, A. Baudisch, F. Colchero, *Science* **376**, 1466 (2022).
2. B. A. Reinke *et al.*, *Science* **376**, 1459 (2022).
3. C. E. Finch, M. C. Pike, M. Witten, *Science* **249**, 902 (1990).
4. C. E. Finch, *J. Gerontol. A Biol. Sci.* **53A**, B235 (1998).
5. C. E. Finch, *Longevity, Senescence, and the Genome* (Univ. Chicago Press, 1990).
6. S. N. Austad, *Methuselah's Zoo: What Nature Can Teach Us About Living Longer, Healthier Lives* (MIT Press, 2022).
7. D. A. Warner, D. A. W. Miller, A. M. Bronikowski, F. J. Janzen, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6502 (2016).
8. J. W. Gibbons, *BioScience* **37**, 262 (1987).
9. H. Cayuela *et al.*, *Biol. J. Linn. Soc.* **128**, 251 (2019).
10. J. D. Congdon, R. D. Nagle, O. M. Kinney, R. C. van Loben Sels, *Exp. Gerontol.* **36**, 813 (2001).
11. A. Bartke, *Rev. Endocr. Metab. Disord.* **22**, 71 (2021).
12. D. E. L. Promislow, M. Tatar, S. Pletcher, J. R. Carey, *J. Evol. Biol.* **12**, 314 (1999).
13. H. Beltrán-Sánchez, C. Finch, *eLife* **7**, e34427 (2018).

10.1126/science.adc9442

ENVIRONMENTAL SCIENCE

The delicate balance of river sediments

Global satellite data quantify changes in sediment flux in 414 rivers

By **Christiane Zarfl**[1] *and* **Frances E. Dunn**[2]

Rivers are more than just streams of water; they also carry sediments from the land, riverbeds, and riverbanks and deposit them downstream or into oceans. Sediments play a pivotal role in defining river morphology on small scales (such as riverine habitats) and large scales (such as river deltas), as well as in shaping river ecosystems by transporting nutrients and pollutants. Most anthropogenic land use often increases sediment erosion and transport, whereas dam building decreases sediment transport. In addition, changes in precipitation frequency and intensity—related to climate change—also affect sediment erosion and discharge dynamics. Data concerning these drivers are often lacking but are essential for effective river basin management. On page 1447 of this issue, Dethier *et al.* (*1*) present an analysis of sediment flux changes for 414 rivers worldwide based on satellite data from 1984 to 2020, evaluated against 130,000 field measurements.

In situ characterization of rivers is often lacking because it is too expensive to sample frequently enough to capture important dynamics across large areas. Remote sensing based on satellite or aircraft data has the potential to make environmental monitoring easier by providing access to observations regardless of remoteness or terrain difficulty. Dethier *et al.* used algorithms to derive fluvial—instream—suspended sediment concentrations from global satellite imagery dating back to the 1980s. Although similar techniques have been used locally before [for example, (*2*)], the scale, both spatial and temporal, and extensive verification of the work by Dethier *et al.* is unprecedented, which allows the most detailed and comprehensive analysis of these trends in the data. Overall, they found a

[1]Center for Applied Geosciences, Eberhard Karls University of Tübingen, Tübingen, Germany. [2]Faculty of Geosciences, Utrecht University, Utrecht, Netherlands.
Email: christiane.zarfl@uni-tuebingen.de; f.e.dunn@uu.nl

global reduction in sediment fluxes of almost 50% in rivers located in the Northern Hemisphere, which was mainly the result of sediment entrapment by dams, but an increase in suspended sediment concentrations of over 40% in rivers located in the Southern Hemisphere, which was driven by deforestation and subsequent mining and agricultural activities.

Earlier studies have hinted at the different facets of these findings. Dams and reservoirs are known to be very efficient retention factors, stopping an estimated 50% of sediments that would otherwise be discharged into the oceans (*3*). Because of an expansion of dams—not only for electricity production but also for flood control, irrigation purposes, or water storage—there has been a global decrease in fluvial sediment fluxes (*4*, *5*). This development is spatially heterogeneous both in trend and quantity. For example, recent modeling of cumulative sediment retention by dams on the Mekong River shows that as little as 4% of the total sediment load is expected to reach the river delta (*6*). At the same time, human activities such as agriculture, land leveling, mining, and infrastructure construction increase erosion rates by orders of magnitude (*7*)—far more than the decrease caused by dams. For example, in the Middle Ages, soil erosion increased by up to a hundredfold after deforestation in Europe, as observed in lake sediment core samples (*8*), and it is estimated that more than 50% of the world's ice-free land area has been altered by 2007 from anthropogenic activities (*9*). Understanding and disentangling the natural and anthropogenic drivers of sediment erosion and the accumulation of sediments remains a highly relevant topic in geologic and environmental sciences (*10*).

Changes in sediment fluxes have huge ramifications for the stability of river systems, their geomorphology, ecology, and the socioeconomic activities that rely on their integrity (see the figure). Increased sediment fluxes can cause deposition in river channels. This has many consequences, including changing species composition and ecosystem productivity by altering habitats, increasing flood risk by reducing channel capacity, and creating economic burdens by requiring dredging, such as in reservoirs to maintain storage capacity (*11*) or in river channels to maintain the navigable depth for ships (*12*). Decreased sediment fluxes in so-called "sediment-starved" water can worsen erosion in river channels, destroying ecosystems and aquatic productivity, potentially leading to changes in food webs, nutrient transfer, and even a reduction in biomass formation and threats to food security of local human populations. Sediment-starved water eroding in river channels can also lead to bank collapse, affecting river engineering and land adjacent to rivers and wrecking geomorphic features such as deltas (*13*) and beaches.

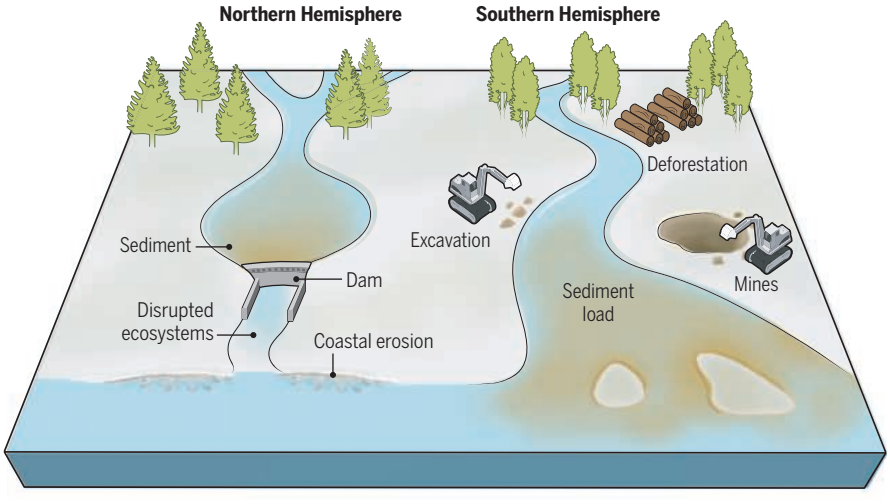
To maintain the physical environment and ensure sustainable use of these systems, these changes must be acknowledged and anticipated to avoid inappropriate decisions that fail to address environmental challenges or that exacerbate unwanted environmental changes. Management decisions must be well informed by data and not by outdated assumptions about the state of river systems, especially given the accelerating rates of change driven by human activities (*14*).

Appropriate management will depend on anticipating future trends, which relies on understanding the key drivers of change. Although projections of fluvial sediment delivery show a continuation of the global decline with some spatial variation (*15*), future data and analyses will contribute to refining projections, particularly for areas of the world where data have been sparse. Historical data can provide information on previously understudied areas and help inform future geoengineering projects based on lessons learned. For example, if fluvial sediment concentrations have recently fallen because of dam construction on the river and more dams are planned, they will not have the same effect as dam construction on a river with no prior dams because sediment concentrations are unable to fall as far. Alternatively, if sediment fluxes have recently increased because of erosion caused by land-use change, the new higher sediment fluxes may be unsustainable and vulnerable to reduction.

Integrated studies that combine new approaches to data mining, such as evaluating satellite data against a set of spatially broad and temporally long monitoring data from the field, can help identify overarching and general patterns in environmental dynamics. Understanding the underlying processes and interdependencies is essential to narrow down the range of potential future developments—that is, to specify scenarios for model projections. This, collectively, is required to provide a sound scientific basis for responding to fluvial sediment flux changes, including the consideration and communication of uncertainties when fostering transparent discussions on sustainable river management. ■

The imbalance of sediment flux between the Northern Hemisphere and Southern Hemisphere

A meta-analysis of 414 major rivers around the world reveals how different human activities affect sediment transport in rivers, and that the Northern and Southern Hemispheres experience problems related to excess retention and erosion of sediment, respectively.



GRAPHIC: N. DESAI/*SCIENCE*

REFERENCES AND NOTES

1. E. N. Dethier *et al.*, *Science* **376**, 1447 (2022).
2. J. M. Martinez, J. L. Guyot, N. Filizola, F. Sondag, *Catena* **79**, 257 (2009).
3. C. J. Vörösmarty *et al.*, *Global Planet. Change* **39**, 169 (2003).
4. J. P. M. Syvitski, C. J. Vörösmarty, A. J. Kettner, P. Green, *Science* **308**, 376 (2005).
5. J. P. M. Syvitski, A. Kettner, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **369**, 957 (2011).
6. G. M. Kondolf *et al.*, *Sci. Total Environ.* **625**, 114 (2018).
7. T. Vanwalleghem *et al.*, *Anthropocene* **17**, 13 (2017).
8. J. A. Dearing, *Clim. Past* **2**, 187 (2006).
9. R. L. Hooke, J. F. Martín Duque, *GSA Today* **22**, 4 (2012).
10. C. Zarfl, A. Lucía, *Curr. Opin. Environ. Sci. Health* **5**, 53 (2018).
11. W. L. Hargrove, D. Johnson, D. Snethen, J. Middendorf, *J. Soil Water Conserv.* **65**, 14A (2010).
12. J. R. Cox, F. E. Dunn, J. H. Nienhuis, M. van der Perk, M. G. Kleinhans, *Anthrop. Coasts* **4**, 251 (2021).
13. J. P. M. Syvitski *et al.*, *Nat. Geosci.* **2**, 681 (2009).
14. F. E. Dunn, P. S. J. Minderhoud, *Commun. Earth Environ.* **3**, 2 (2022).
15. F. E. Dunn *et al.*, *Environ. Res. Lett.* **14**, 084034 (2019).

10.1126/science.abq6986

WRITING STUDIES

Alphabetized

The origins of the world's chief writing system come to life in an illuminating historiography

By **Andrew Robinson**

Characters inscribed on the Serabit el-Khadim sphinx (1800–1500 BCE) may be a prototype of the alphabet.

PHOTO: THE TRUSTEES OF THE BRITISH MUSEUM/CC BY-NC-SA

The alphabet is the chief writing system in use today. Not surprisingly, there have been many notable books about its history. In the 20th century, for example, there was David Diringer's *The Alphabet* (1948), Joseph Naveh's *The Early History of the Alphabet* (1982), and John F. Healey's *The Early Alphabet* (1990). Johanna Drucker's new book, *Inventing the Alphabet: The Origins of Letters from Antiquity to the Present*, is not another history—it is a historiography, addressing the intellectual history of this crucial topic for the first time.

Drucker presents and analyzes the work of alphabet scholars from the time of Herodotus and Socrates in the fifth century BCE; through the long-dominant biblical account, which centered on Moses and the stone tablets on which the Ten Commandments were mysteriously written; up to the current debates surrounding 20th-century archaeological finds of a few hundred very early alphabetic inscriptions in Egypt and Palestine. "Who knew what when about the alphabet?" she asks. "And how did the way they knew it—through texts, images, inscriptions, or artifacts—affect their conception of the identity and origin of alphabetic writing?"

Drucker is a professor of bibliographical studies and has been a designer of printed words and letters for almost half a century. Her scholarship and aesthetics make for an illuminating, if undoubtedly academic, study, illustrated with historic documents, archaeological finds, drawings of inscriptions, and copious tables of recondite letters compiled by generations of scholars. Regrettably, no maps are presented to show the usage of the alphabet, and the occasional surprising error appears, such as her description of the bilingual Rosetta stone as a "multilingual" inscription.

The ancient Greek historian Herodotus is a key source cited in the book. In a widely accepted translation of *The Histories*, he famously wrote of the origin of the alphabet in Greece as follows: "These Phoenicians…brought with them to [Greece], among many other kinds of learning, the alphabet, which had been unknown before this…As time went on the sound and the form of the letters were changed…In so doing, [the Greeks] gave to these characters the name of *Phoenician*."

However, in the original Greek, Herodotus writes "phoenikeia grammata," meaning "Phoenician letters," and does not use the word "alphabet." The Latin word "alphabetum"—derived from the first two Greek letters alpha and beta—is not documented until centuries after the time of Herodotus, notes Drucker.

Such uncertainty has created persistent and continuing arguments among scholars. Classicists tend to cling to the long-held view that the Greeks invented the alphabet. By way of example, Drucker quotes Roger Woodard in his *Greek Writing from Knossos to Homer* (1997): "Of the many splendid achievements of the ancient Greeks, the alphabet was perhaps the most marvelous and certainly the most influential." She comments: "In this version, instead of being an adaptation and modification, the alphabet becomes a unique invention, even if, as Woodard goes on to say, 'it stands on the shoulders of the consonantal script of the Phoenicians.'"

Inventing the Alphabet
Johanna Drucker
University of Chicago Press, 2022. 384 pp.

Drucker herself tends to the more widely held scholarly view that Greeks living in Phoenicia invented their alphabet, inspired by the Phoenician script, and that from there, it spread to the mother country during the first half of the first millennium BCE. She also presents the limited but thought-provoking evidence from Sinai and Egypt that the Phoenicians may have been inspired to create the alphabet by others.

A small sphinx discovered in 1905 in Serabit el-Khadim, Sinai, by Flinders Petrie and currently housed in the British Museum offers support for this theory. Inscribed in both Egyptian hieroglyphs and "Proto-Sinaitic" signs that resemble certain of the hieroglyphs, the signs may be a prototype of the alphabet. The even earlier Wadi el-Hol inscriptions, discovered in the 1990s by John and Deborah Darnell, support this deduction. Such findings suggest that the first alphabet was inspired by the script of the ancient Egyptians, perhaps as early as 1900 BCE. And yet, as always seems to be the case with the origins of the alphabet, this evidence is insufficient and awaits the discovery of more inscriptions, no doubt followed by more arguments among scholars. ■

10.1126/science.abo7809

The reviewer is the author of *The Story of Writing: Alphabets, Hieroglyphs and Pictograms* (Thames & Hudson, 2007). Email: andrew@andrew-robinson.org

FOREST ECOLOGY

Dispatches from the redwood rebellion

Tourism fails to compensate for lost timber jobs in protected forests, leading some to theft

By **Jeremy B. Yoder**

The concept of stealing a tree seems, at first, like a category error. Under the spreading boughs of a centuries-old valley oak or amid the green-lit colonnade of a basswood forest, a tree theft seems as likely as the theft of a mountain or a river. Of course, we do take trees out of the forest for fuel and timber and fiber, just as we exploit mountains for mining and reroute rivers for irrigation. And if trees are a resource with monetary value, then it follows that they can be stolen.

In *Tree Thieves*, environmental journalist Lyndsie Bourgon examines tree theft throughout forestry history and discusses the modern global trade in stolen timber. The heart of the book is devoted to individuals who navigate the boundary between legal lumber harvest and tree poaching in California's Redwood National and State Parks.

The Redwood National and State Parks were first established in the early 20th century to protect old-growth coast redwoods (*Sequoia sempervirens*). These trees can live more than 2000 years, growing into enormous pillars of fine, durable wood, taller than any other tree species. A grove of old-growth redwoods is a living cathedral, valued for its beauty and for the foundational role these trees play in the ecosystems they inhabit. It is also worth a fortune as timber. That dual status means that efforts to protect redwoods are often in direct conflict with the needs of communities that depend on them for income.

The addition of Redwood National Park to an existing collection of state parks in 1968 was supposed to bring employment and tourist dollars to replace timber jobs. However, perhaps inevitably, it fell far short of the economic support provided by the midcentury lumber industry. In Orick, California, a community built by timber companies and spiraling into decline after their retreat, Bourgon follows a network of men trying to extract a living from the forest in their backyards. Known locally as "the Outlaws," they poach redwood in chunks chainsawed from trees, fallen and standing.

By their own account, the Outlaws' timber theft is part desperation and part protest. There are few paying jobs left in Orick, and locals who remember when it was possible to make a living from the forest see the national park as an obstruction imposed by distant, urbanite environmentalists.

Tree Thieves places the Outlaws' actions in broader context, providing a pocket history of forest regulation in English law that explains how public rights to forest resources were guaranteed from the time of the Magna Carta and how the privatization of forests by the wealthy has been a continual source of discontent for those without privilege. Errol Flynn's Robin Hood delivering one of "the king's deer" to Prince John's banquet table is an echo of such struggles.

Bourgon also takes in a global view of the trade in poached lumber, worth up to $157 billion annually, as much as 30% of worldwide wood sales. Along the way, readers meet law enforcement officers and scientists working to detect and deter tree theft.

Spotting illegally taken wood amid shiploads of lumber is a challenge, and linking a piece of poached wood to its origin is even more difficult, despite the availability of genetic testing. Bourgon describes how the prosecution of one of the Orick Outlaws relies on rangers finding blocks of redwood that slot neatly into a hole in a damaged tree.

Tree Thieves glimpses alternatives to the dynamics at play among the California redwoods. Bourgon profiles community-managed forests near her home in British Columbia and tours conservation concessions managed by towns and Indigenous communities in the Peruvian Amazon. These schemes can give locals—in some cases people who lived on the land before colonization—real agency in forest conservation. But they have their own shortcomings. In the Amazon, for example, locals lament the lack of governmental support for the prevention of incursions by clear-cutting squatters. In British Columbia, the Sunshine Coast Community Forest loses old-growth Douglas firs to poaching even as its managers issue hundreds of free permits for sustainable harvest.

The book's unavoidable conclusion is that the problem manifest in timber poaching is not the destruction of a particular tree or the failure of a conservation plan but rather a social and economic system that roots personal identity in wage-earning work (or lack thereof) and that describes a tree by its value as board feet in a lumberyard. *Tree Thieves* thus suggests that the theft of a tree may be a category error after all—but it is an error so pervasive that we don't know we're standing in its shade. ■

10.1126/science.abq6316

Tree Thieves: Crime and Survival in North America's Woods
Lyndsie Bourgon
Little, Brown Spark, 2022. 288 pp.

The reviewer is at the Department of Biology, California State University Northridge, Northridge, CA 91330, USA. Email: jeremy.yoder@csun.edu

Maize grows in Oaxaca, Mexico, in 2018.

PHOTO: JON G. FULLER, JR./VWPICS/AP IMAGES

PODCAST

Fresh Banana Leaves: Healing Indigenous Landscapes Through Indigenous Science
Jessica Hernandez
North Atlantic Books, 2022. 256 pp.

Indigenous communities have long recognized the importance of holistic land management and environmental stewardship, but these practices are frequently undervalued in Western conservation policies. This week on the *Science* podcast, Jessica Hernandez reveals how Indigenous Latin American environmental knowledge can help foster more sustainable agricultural systems.
https://bit.ly/3xmfXrd

10.1126/science.add0302

LETTERS

A proposed law would allow road construction in Iguaçu National Park, the largest remnant of the Atlantic Forest biome of southern Brazil.

PHOTO: JOSE SABINO

Edited by **Jennifer Sills**

A wrong turn on roads in Brazil's national parks

The Brazilian Senate is debating a proposed law to reopen the Colono road crossing Iguaçu Falls National Park (*1*). This road was closed 20 years ago because it violated reserve management rules (*2*). National parks ensure full environmental protection with no human interference (*3*). Reopening the road would undermine the park's purpose of conserving the last remnant native vegetation in the southwestern Atlantic Forest (*2*). To circumvent this, the proposal suggests the creation of a protected area within the park that has fewer conservation requirements and would allow road construction. Changing the law on these terms would serve as a concerning precedent for other reserves in this megadiverse country.

Land clearing for economic development resulted in the near-complete clearing of Atlantic Forest in the Iguaçu Falls region during the 20th century (*4*). A paved road will escalate edge effects, with negative consequences for local climate, biodiversity, and ecosystem services (*5*). The road will also increase the invasion of alien species and the scale of illegal activities such as poaching and logging (*6*). Threatened species, such as the jaguar (*Panthera onca*), rely on the national park as a last refuge in the western Atlantic Forest (*7*). In addition, traffic will cause chemical and acoustic pollution as well as pose a threat to wildlife (*8*). About 475 million wild vertebrates already lose their lives annually in Brazil as a result of collisions with motor vehicles (*9*).

Those supporting the road claim that it would improve the local economy by increasing tourism, but there is little evidence that it would have this effect. Economic activities should be restricted to the already active buffer zone. The local economy can be stimulated by supporting conservation tourism, which could include improving infrastructure and environmental education, creating visitor centers, regulating activities around the park, and training more personnel to receive visitors. This strategy would minimize damages to the core zone and make it unnecessary to legalize the illegal.

Lucas Rodriguez Forti[1], Rodrigo Lingnau[2], Judit K. Szabo[1,3]*

[1]Universidade Federal da Bahia, Salvador, Brazil. [2]Universidade Tecnológica Federal do Paraná, Francisco Beltrão, Brazil. [3]College of Engineering, IT, and Environment, Charles Darwin University, Darwin, NT 0909, Australia.
*Corresponding author.
Email: judit.szabo@cdu.edu.au

REFERENCES AND NOTES

1. Brazil, Projeto de Lei da Câmara no 61, de 2013 (2022) [in Portuguese].
2. Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio), "Plano de Manejo do Parque Nacional do Iguaçu" (2018) [in Portuguese].
3. International Union for Conservation of Nature, Protected Areas, Category II: National Park; www.iucn.org/theme/protected-areas/about/protected-areas-categories/category-ii-national-park.
4. M. B. Carlucci, V. Marcilio-Silva, J. M. Domingues Torezan, in *The Atlantic Forest*, M. C. M. Marques, C. E. V. Grelle, Eds. (Springer Cham, 2021), pp. 91–111.
5. W. F. Laurance *et al.*, *Trends Ecol. Evol.* **24**, 659 (2009).
6. W. F. Laurance *et al.*, *Nature* **513**, 229 (2014).
7. ICMBio, "Plano de Ação Nacional para a conservação da onça-pintada" (2013) [in Portuguese].
8. J. Khan *et al.*, *Sci. Tot. Environ.* **634**, 661 (2018).
9. Sistema Urubu, Estradas e Animais (2022); https://sistemaurubu.com.br/dados/ [in Portuguese].

10.1126/science.add2508

Diplomacy for the world's hottest sea

Countries that border the Persian or Arabian Gulf [hereafter, "Gulf" (*1*)] are adopting ambitious global commitments to protect their marine environments. The United Arab Emirates (UAE) has committed to protecting 30% of its lands and waters by 2030 (*2*), with other Gulf states expected to soon join. Gulf countries will not meaningfully meet these and other global commitments (such as the post-2020 biodiversity framework) without rebuilding trust, exchanging knowledge, and jump-starting conservation coordination across their maritime boundaries. Gulf scientists have an overlooked role in this work.

The Gulf's political boundaries include Bahrain, Iran, Iraq, Kuwait, Oman, Qatar, Saudi Arabia, and the UAE, making the regional networking of marine protected areas (MPAs) critical to system-wide conservation success (*3*). During the summer, the Gulf is the world's hottest marine system, with many organisms living near their physiological limits (*4*, *5*). The Gulf is a global hot spot for coastal development and desalination, causing further declines to coral reefs and sea grass meadows (*6*–*8*), as well as direct threats to dugongs and other transboundary species.

In contrast to remote and largely uninhabited ecosystems like Antarctica, highly degraded ecosystems are overlooked settings for advancing science diplomacy (*9*, *10*). The softening of political tensions between many Gulf countries opens possibilities for action. Regional funding, coordinated through the Gulf Cooperation Council as well as national funding bodies,

should be allocated to help collaborating scientists pursue networked Gulf studies. Gulf-wide peer networks among scientists should be established to build trust through data sharing, strengthening relationships for "data diplomacy" (*11*). In addition, scientists should collaborate with capable boundary organizations that can facilitate the planning of future transboundary MPAs, an intervention critically needed in the Gulf.

Scientists and international institutions do not depend on formal diplomatic channels to act (*12*). Starting now will create a foundation for coproducing state-led ocean science diplomacy in the coming decades, ushering in a new era of peace and ocean protection in the region.

Nadia Al-Mudaffar Fawzi[1], Clare M. Fieseler[2]*, Brian Helmuth[3], Alexandra Leitão[4], Mehsin Al-Ainsi[5], Mohammad Al Mukaimi[6], Mohammad Al-Saidi[7], Fahad Al Senafi[6], Ivonne Bejarano[8], Radhouan Ben-Hamadou[4,5], Josh D'Addario[9], Ahmad Mujthaba Dheen Mohamed[5], Bruno W. Giraldes[4], Lyle Glowka[10], Maggie D. Johnson[11], Brett P. Lyons[12], Daniel Mateos-Molina[13], Christopher D. Marshall[14,15], Sayeed Mohammed[16], Pedro Range[4], Mohammad Reza Shokri[17], John M. K. Wong[18], Nicholas D. Pyenson[2]

[1]Marine Science Centre, University of Basrah, Basra, Iraq. [2]National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA. [3]Coastal Sustainability Institute and School of Public Policy and Urban Affairs, Northeastern University, Boston, MA 02215, USA. [4]Environmental Science Center, Qatar University, Doha, Qatar. [5]Department of Biological and Environmental Sciences, College of Arts and Sciences, Qatar University, Doha, Qatar. [6]Department of Marine Science, College of Science, Kuwait University, Kuwait City, Kuwait. [7]Department of International Affairs, Qatar University, Doha, Qatar. [8]Department of Biology, Chemistry, and Environmental Sciences, American University of Sharjah, Sharjah, United Arab Emirates. [9]The Open Data Institute, London, UK. [10]Biodiversity Strategies International, Abu Dhabi, United Arab Emirates. [11]Red Sea Research Center, Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia. [12]Weymouth Laboratory, Centre for Environment, Fisheries and Aquaculture Science, Weymouth, UK. [13]Emirates Nature World Wildlife Fund, Dubai, United Arab Emirates. [14]Department of Marine Biology, Texas A&M University, Galveston Campus, Galveston, TX 77554, USA. [15]Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX 77843, USA. [16]Arab Climate Young Movement Qatar, Doha, Qatar. [17]Department of Animal Sciences and Marine Biology, Faculty of Life Science and Biotechnology, Shahid Beheshti University, Tehran, Iran. [18]Aquatic Research Center, Ministry of Environment and Climate Change, Doha, Qatar.
*Corresponding author. Email: fieselerc@si.edu

REFERENCES AND NOTES

1. The name of this body of water remains contentious. We use "Gulf" to omit any geographic descriptors that detract from a productive dialogue.
2. High Ambition Coalition, HAC Member Countries (2021); www.hacfornatureandpeople.org/hac-members.
3. S. E. Schill *et al.*, *PLOS ONE* **10**, 0144199 (2015).
4. B. M. Riegl *et al.*, *Coral Reefs of the Gulf: Adaptation to Climatic Extremes*, B. M. Riegl, S. J. Purkis, Eds. (Springer, Netherlands, 2012).
5. C. C. Hume *et al.*, *Scientif. Rep.* **5**, 8562 (2015).
6. J. Burt, *City* **18**, 760 (2014).
7. P. F. Sale *et al.*, *Ambio* **40**, 4 (2010).
8. N. J. Edmonds *et al.*, *Mar. Pollut. Bull.* **163**, 111915 (2021).
9. C. M. Brooks *et al.*, *Science* **354**, 185 (2016).
10. P. A. Berkman *et al.*, *Science* **358**, 596 (2017).
11. A. Boyd *et al.*, *Sci. Diplom.* **8**, 1 (2019).
12. K. M. Lord, V. C. Turkekin, *Science* **315**, 769 (2007).

10.1126/science.add1555

Minimize food loss and waste to prevent crises

Russia's invasion of Ukraine and the associated economic and trade consequences have added to the global food security disruptions caused by the COVID-19 pandemic (*1*). Russia produces 6 to 20% of the world's potash, phosphate, and ammonia, and the trade embargo implemented in response to the invasion has led to increased demand for these important sources of fertilizer macronutrients (*2*). In addition, Russia and Ukraine together account for 18 to 70% of global wheat, maize, and sunflower exports (*3*), and Ukraine maintains a large seed bank that is now at risk of destruction (*4*). The fall in exports has led to food shortages and increased prices, including maize and wheat price hikes that are, respectively, 55% and 91% higher than prices for those products in 2021 (*5*). In the face of these challenges, strategies to ensure food security should be explored, including the prevention of food loss and waste.

More than 931 million tons of food were wasted in 2019 (*6*). About 8 to 10% of total global greenhouse gas emissions are associated with food loss and waste, with a monetary value of 1 trillion USD per year (*7*). Tackling the food loss and waste challenge would ensure a more sustainable, and reliable, long-term food supply chain (*8*).

During their May meeting, the G7 foreign ministers committed to reducing food loss and waste to enhance food security and fulfill the Sustainable Development Goals (*9*). The United States has identified four focus areas: preventing food waste, increasing food recovery, recycling food waste, and helping to coordinate food waste reduction. These efforts have been recommended to the US Congress to tackle food waste in the 2023 Farm Bill (*10*).

Food loss and waste account for 17% of total global food production (*6*). A coordinated global effort to reduce this wasted food could result in substantial food savings, mitigating the uncertainties caused by crises like the pandemic and Russia's invasion. Intergovernmental collaboration, in the form of organizations like the Intergovernmental Platform on Biodiversity and Ecosystem Services and the Codex Alimentarius Commission, will be essential to expedite the formulation of standards and codes of practice for this effort (*11*, *12*).

Siming You[1]*, Christian Sonne[2], Roger Ruan[3], Peng Jiang[4]

[1]University of Glasgow, Glasgow, UK. [2]Aarhus University, Roskilde, Denmark. [3]University of Minnesota, St. Paul, MN 55455, USA. [4]Sichuan University, Chengdu, China.
*Corresponding author.
Email: siming.you@glasgow.ac.uk

REFERENCES AND NOTES

1. D. Laborde *et al.*, *Science* **369**, 500 (2020).
2. G. Ibendahl, "The Russia-Ukraine conflict and the effect on fertilizer" (2022); https://agmanager.info/sites/default/files/pdf/Ibendah_Fertizer_RussiaUkraine_03-08-22.pdf.
3. C. Hanson, "The Ukraine crisis threatens a sustainable food future," GreenBiz (2022).
4. M. Angel, "Ukraine's giant seed bank at risk of being lost as war rages," *Reuters* (2022).
5. The World Bank, "World Bank commodities price data" (2022); https://thedocs.worldbank.org/en/doc/5d903e848db1d1b83e0ec8f744e55570-0350012021/related/CMO-Pink-Sheet-May-2022.pdf.
6. UN Environment Programme, "UNEP food waste index report 2021" (2021).
7. C. Mbow *et al.*, in "Special report—climate change and land" (Intergovernmental Panel on Climate Change, 2019), pp. 437–550.
8. I. Delabre *et al.*, *Sci. Adv.* **7**, eabc8259 (2021).
9. UK Foreign, Commonwealth, & Development Office, "Statement by foreign ministers of G7 countries and the High Representative of the European Union" (2022).
10. E. M. Broad Lieb *et al.*, "Opportunities to reduce food waste in the 2023 Farm Bill" (2022).
11. E. Turnhout *et al. Science* **373**, 1093 (2021).
12. J. Kluytmans, *Nature* **501**, 316 (2013).

10.1126/science.add2008

TECHNICAL COMMENT ABSTRACTS

Comment on "The early origin of a birdlike inner ear and the evolution of dinosaurian movement and vocalization"

Romain David, Mario Bronzati, Roger B. J. Benson

Hanson *et al*. (Research Articles, 7 May 2021, p. 601) claim that the shape of the vestibular apparatus reflects the evolution of reptilian locomotion. Using biomechanics, we demonstrate that semicircular canal shape is a dubious predictor of semicircular duct function. Additionally, we show that the inference methods used by Hanson *et al*. largely overestimate relationships between semicircular canal shape and locomotion.

Full text: dx.doi.org/10.1126/science.abl6710

Response to Comment on "The early origin of a birdlike inner ear and the evolution of dinosaurian movement and vocalization"

Michael Hanson, Eva A. Hoffman, Mark A. Norell, Bhart-Anjan S. Bhullar

David *et al*. claim that vestibular shape does not reflect function and that we did not use phylogenetic inference methods in our primary analyses. We show that their claims are countered by comparative and direct experimental evidence from across Vertebrata and that their models are empirically unverified. We did use phylogenetic methods to test our hypotheses. Moreover, their phylogenetic correction attempts are methodologically inappropriate.

Full text: dx.doi.org/10.1126/science.abl8181

NEWS
Can biofuels really fly? p. 1394

POLICY FORUMS
Strengthen climate adaptation research globally p. 1398
How trade policy can support the climate agenda p. 1401

REVIEWS
Current global efforts are insufficient to limit warming to 1.5°C p. 1404
Harnessing the potential of nature-based solutions for mitigating and adapting to climate change p. 1410
Climate change and the urgency to transform food systems p. 1416
Getting ahead of climate change for ecological adaptation and resilience p. 1421

RELATED ITEMS
EDITORIAL p. 1361 NEWS STORY p. 1372 PODCAST

TIME TO ACT

By **Caroline Ash, Bianca Lopez, Jesse Smith, Sacha Vignieri,** *and* **Brad Wible**

Most people now fully accept that climate change is occurring and is caused by human activities. In light of current public attitudes, one could be forgiven for thinking that the data that support these conclusions are relatively new. The truth, however, is that climate scientists have known and warned for decades that our activities are leading to dangerous climate change, the effects of which we are experiencing now. Technologies for alternative energy sources have also existed for decades, yet political and financial interests have prevented their widespread uptake, as well as the transformational economic and social change needed to end our alteration of climate. Today, when our options are limited and our need is urgent, these same forces are preventing transformation. Even if carbon emissions are halted today, the climate will continue to warm, with profound impacts on the Earth system. In this special issue, we explore ways that science can help guide us to a more promising climate future, from understanding where we are in our climate trajectory and how natural systems may respond, to providing options for mitigating climate change and adapting our systems (and ourselves) to the forces we have unleashed.

ILLUSTRATION: MYRIAM WARES

Carbon dioxide emissions from fossil fuel use—along with landscape destruction and biodiversity loss—are altering our planetary system. We can still prevent the most catastrophic changes, but time is running out.

Sandbags have been placed to protect the coast in Temwaiku, South Tarawa, Kiribati. Holding the line against climate hazards is often vital for some territories but may be maladaptive over the long run.

PHOTO: KADIR VAN LOHUIZEN/NOOR/REDUX

POLICY FORUM

Strengthen climate adaptation research globally

More international incentives and coordination are needed

By **Alexandre K. Magnan[1,2], Ariadna Anisimov[1], Virginie K. E. Duvat[2]**

The window of opportunity for climate adaptation action is closing fast because of warming and development trends (*1*). Yet efforts toward enhancing adaptation policy, implementation, and finance are not considered at adequate scale globally (*2, 3*). Scaling up such efforts requires substantial international investment, which we argue should include support for transdisciplinary adaptation research to enhance scientific foundations to feed into more effective policy engagement and funding for implementation. Here, we identify opportunities associated with three scientific frontiers: understand the potential for effective climate risk reduction (including understanding maladaptation, residual risk, and adaptation limits); assess systemic, cascading, and transboundary risks; and track adaptation progress.

The recent report of Working Group II (WGII) of the Intergovernmental Panel on Climate Change (IPCC) shows that even under a low emissions scenario aligning with the +1.5°C/+2°C temperature target, the world is on track to experience severe climate risks before the end of this century (*1*, *4*). It also reports widespread and substantial risk levels occurring at lower global warming levels than in previous assessments (*4*). In addition, it is estimated that by 2100, global climate risk will increase by two- to fourfold under 2°C and 4°C of global warming, respectively (*5*), and that half of humankind is living in areas that are highly vulnerable to climate change (*1*). These conclusions make the case that adaptation is no longer an option but an imperative for socioecological systems over this century and beyond.

Another element emphasized by the report is that besides being a national and local concern, adaptation is also a global responsibility. The United Nations Framework Convention on Climate Change (UNFCCC), the Adaption Fund, the IPCC report, and science and policy communities more broadly increasingly recognize that in many cases, localized adaptation decisions can have implications on distant places through, for example, supply chains, markets, or movements of people. Thus, implementing sound climate adaptation should ultimately be seen as a shared responsibility, both nationally and internationally, toward long-term socioeconomic well-being and equity.

One question therefore is how science can help address the adaptation challenge at multiple scales. Prior calls for adaptation research have emphasized the need to support improved knowledge on decision-making, vulnerabilities, forecasting, and action (*6*). Although these pleas contributed to boosting adaptation studies and raising the policy profile of the topic (in international negotiations, development assistance, and national to local policies), there is a need, we argue, to move a step further—first, because policies are not yet equipped to deal with emerging issues emphasized by the WGII report on understanding adaptation effectiveness, the risk of maladaptation, residual risks, adaptation limits, and compound and cascading risks, and second, because increasing and accelerating climate risks call for scaling up adaptation knowledge and policy impact. We thus need both more research and more coordination internationally.

UNDERSTAND EFFECTIVE CLIMATE RISK REDUCTION

The IPCC WGII report reinforces conclusions drawn from the 2019 Special Reports on Land and the Ocean and Cryosphere that societal adaptation can substantially decrease climate risk under all warming scenarios (*5*). Understanding the effectiveness of adaptation to reduce current and future climate risk is critical for at least two reasons: (i) assessing whether current adaptation efforts are sufficient or not in a context of increasing warming and (ii) identifying the room to maneuver in terms of risk reduction—for example, when comparing risk levels under low- and high-adaptation scenarios. Such knowledge is decisive to help set up adaptation policy targets at a given scale and hence structure and galvanize action across systems, stakeholders, and scales.

The IPCC WGII, however, raises three major concerns. First, the scientific literature still provides little evidence on effective risk reduction associated with the adaptation-related responses that are reported on the ground (*2*). This is partly inherent in risk reduction often needing time to become (or not) evident and measurable. Yet although there is knowledge on the observed outcomes of a wide range of actions that aim to reduce risk, assessing these outcomes in the future and under changing climate conditions remains difficult. The development of robust adaptation policies, however, requires such a forward-looking perspective.

Second, there are concerns about the increase of the risk of maladaptation, when measures implemented in the name of adaptation reveal counterproductive effects that lead to increasing long-term exposure and vulnerability to climate change. For example, engineered structures such as seawalls used

[1]Institute for Sustainable Development and International Relations (IDDRI-Sciences Po), Paris, France. [2]La Rochelle University & CNRS LIENSs, La Rochelle, France. Email: alexandre.magnan@iddri.org

to protect coasts reduce risks in the short term but can increase socioecological systems' exposure and vulnerability over time through, for example, the degradation of coastal environments (such as beach loss), the densification of built assets in newly "protected" areas, and the (false) idea that populations that live behind such structures are safe from coastal hazards. Lock-in effects are at work that insidiously reinforce climate risks in the medium to long term, and counteracting these processes must fully be part of any adaptation strategy.

Third, emerging forward-looking analyses involving various methods (climate modeling, socioeconomic projections, and expert judgments) and a mix of quantitative and qualitative findings suggest that ambitious adaptation cannot fully prevent a part of climate risk from manifesting, even at low levels of warming for some high-risk regions (*5*, *7*). This refers to the likely existence of residual risks—risks that remain despite adaptation—and adaptation limits, both topics that remain under-investigated within adaptation research (*2*).

As a result, policy-makers and practitioners still face major knowledge gaps on the context-specific potential effectiveness of a range of adaptation-related measures to reduce climate risk over the long run (*1*, *6*), including their potential counterproductive effects (maladaptation), the conditions under which adaptation will not be possible at all (adaptation limits), and the potential for unavoidable consequences (residual risks). One way to speed up things would be for the UNFCCC to call for the development of an IPCC Special Report on "Evidence of and enabling conditions for effective climate adaptation across scales" within the future Seventh Assessment Cycle (AR7). Concretely, that would mean a coalition of Parties to the UNFCCC and/or a formal decision at the 28th Conference of the Parties (COP28)—in 2023 and at the time of the first Global Stocktake on climate action—to put this proposal on the agenda of the scoping meeting of the AR7.

As shown with the IPCC 1.5°C report (*8*), Special Reports could serve as catalysts to raise awareness among policy-makers and funders about frontline research topics, accelerate the production of knowledge, and synthesize scattered scientific information. Our proposed AR7 Special Report could play such a role, for example, by investigating a series of real-world cases across regions and sectors, and/or undertaking assessments for representative territorial archetypes (such as middle-size cities, high mountains, urban atoll islands, and small farming areas) to foster a system-level understanding of effective adaptation.

Besides calling for more applied transdisciplinary research, we also call for a higher level of structuring of scientific efforts to ensure that consistent information is available at the global level. Although several scientific methods have started to emerge on assessing adaptation potential effectiveness, the risk of maladaptation, and adaptation limits, they are used on an individual study basis and are not applied to a wide diversity of sociogeographical contexts. Scaling up such efforts requires more coordination across methods and study contexts, together with a systematic stocktake of scientific advancements. This is definitely not a call for homogenizing scientific methods and neglecting the benefits of a diversity of approaches but for ensuring that multiple sources of scientific information can be brought together to support, for example, IPCC authors in their literature review effort. Yet an international platform on adaptation science is needed to do such coordination and, in collaboration with other institutions, forge closer ties at the science-policy interface.

Because setting up such a platform could take a long time and goes beyond the mandate of the IPCC, we do not advocate for inventing a new organization or institution but rather for building on existing initiatives such as the non-UN Global Adaptation Mapping Initiative or the UN World Adaptation Science Program, which are both in close relation with the IPCC. These initiatives started to organize knowledge (such as systematic literature reviews) and the research community (such as the biannual Adaptation Futures conference) but remain limited in scope and therefore in their ability to provide a singular platform for adaptation data, methods, case studies, and findings on what works or could work as well as under which conditions (such as social, economic, institutional, and environmental).

The associated near-term investment in time, money, and people is competing with the need for actual implementation of adaptation projects, but we argue that these costs will be vastly outweighed by the benefits that accrue over the long run. One example is how such a scientific platform could support UNFCCC policy discussions on Loss and Damage on a more regular basis and with a higher granularity than can the IPCC. Loss and Damage is a highly sensitive topic that touches on residual risks resulting from adaptation efforts that are not effective enough and from the reaching of adaptation limits, together with contentious issues about the attribution of observed impacts to climate change versus nonclimate drivers (*9*). Both increased and coordinated scientific information on various types of losses could be integral (i) to help structure scientifically informed international discussions based on ground-rooted data and (ii) to support the Santiago work program—established by UNFCCC in November 2021 to provide technical support, especially to developing countries—by either matching financial instruments with different types of losses and damages (such as noneconomic) or by developing relevant disbursement criteria. In the end, it will help buy time by targeting more effective action.

ASSESS CASCADING, COMPOUNDING, AND TRANSBOUNDARY RISKS

The IPCC WGII also emphasizes the need to consider cascading impacts; compounding risks (cumulative interactions between several risks and/or risk drivers); and transboundary risks across sectors, jurisdictions, and population groups, both within and across borders (*10*). These processes will combine to generate snowball effects and in turn substantially influence the magnitude, life span, rate of emergence, and spatial spreading of individual risks across systems (*4*). This will likely lead to severe climate risks being higher, lasting longer, and occurring both sooner and at larger scales than assessed in the IPCC WGII report. Another challenge therefore touches on the need for adaptation policies to be better prepared to anticipate and manage systemic risks. In that respect, the UNFCCC could encourage Parties to systematically include a new component or chapter into their official adaptation documents (National Adaptation Plans and Adaptation Communications) to describe the cascading, compounding, and transboundary risks that they are concerned with, as well as strategies to tackle them.

In practice, such a proposal raises numerous questions—for example, on Parties' willingness to consider complex and nondomestic risk issues or their capacity to undertake additional work while already burdened by UNFCCC reporting processes. However, early signs do exist of better policy consideration of cascading, compounding, and transboundary risks—for example, in the UK Climate Change Committee's independent risk assessment in 2021. Analytical frameworks and tools are also emerging on the physical, ecological, and human cascading and compounding consequences of climate change (including environmental and societal tipping points). Expert judgment methods, for example, are used to map risk interactions, often in a rather qualitative way (*4*). But as for emerging studies on assessment of potential effectiveness, maladaptation, limits, and residual risks, these remain topic-limited (such as on agricultural commodities and international trade) and lack higher-level structuring in terms of

method and sharing of findings, as well as policy impact. The Adaptation Without Borders Initiative offers an opportunity toward such a scaling up and structuring process; it started to highlight cases of transboundary risks and governance (including assessment methods) and to engage with policy partners from UNFCCC and several regions (such as Hindu Kush Himalaya).

TRACK ADAPTATION PROGRESS

The IPCC WGII undertook a monumental review of the scientific literature (*2*, *4*) to emphasize, for example, that adaptation-related responses are on the rise worldwide and are rather behavioral (for example, improvement of homes and changes in crops) than technical or institutional. In parallel, during COP26 in Glasgow in 2021 the international climate policy community called for enhancing "understanding of [...] the methodologies, indicators, data and metrics, needs and support needed for assessing progress towards [adaptation globally]" (*11*). A global picture of adaptation efforts is foundational to inform the 5-year Global Stocktake cycle under the UNFCCC through answering two important questions: (i) whether we are collectively on track to adapt to, for example, a 1.5°C world (ii) and how to define more precise global-level adaptation targets that will drive, as seen for mitigation, more ambitious national adaptation policies in a way that they are also consistent with a global perspective.

To date, however, no formal, internationally agreed-on process is in place to assess adaptation progress globally and as comprehensively as possible (*12*). Policy-oriented initiatives exist—such as the Adaptation Gap Report (AGR) of the United Nations Environment Program (UNEP) (*3*)—among others, that deliver regular assessments of global adaptation gaps and progress. However, these assessments usually rely on a limited type of information (for example, mainly from policy documents and international donors in AGR), which makes their findings relatively specific in terms of the adaptation dimension(s) they consider (for example, national policy in AGR). A more comprehensive understanding of adaptation is necessary, which calls for gathering information beyond quantitative metrics, beyond policy and scientific documents, and at multiple scales. To this end, a scientific international initiative to track adaptation progress globally is needed that matches several criteria: be based on innovative scientific methods (such as a combination of modeling and expert judgments) and procedures (peer reviewing); overcome the usual barriers of quantitative data availability (*6*) by bringing together multiple sources of information (also qualitative, gray literature, and traditional knowledge); be distinct from the UNFCCC process to ensure independent outcomes; and at the same time, be policy-relevant enough to feed the UNFCCC discussions and stocktake(s). Some initiatives are emerging, such as the Global Adaptation Progress Tracker (GAP-Track), that offer potential relevant scientific architectures, provided that their scope is broadened (for example, for the GAP-Track, from a coastal focus to a wider spectrum of human settlements such as cities, rural systems, and mountains).

FINANCING ADAPTATION SCIENCE

The gap is widening between the estimated adaptation needs and documented finance allocated to adaptation (*3*). Whereas it is clear that more funding for adaptation action on the ground should remain a global priority (*11*), we argue for enlarging the scope of "climate finance"—originally centered around providing support to developing countries—to accelerate transdisciplinary research. Social sciences in particular need to be better integrated into climate science, which being historically focused on physical and natural sciences provides limited insights on critical dimensions of adaptation (*6*), such as risk perceptions, the societal acceptability of policies and measures, and the role of socioeconomic equity. Our call for more research and more coordination internationally can help.

> **"...implementing sound climate adaptation [is] a shared responsibility, both nationally and internationally..."**

First, a strong research component should be systematically included into each adaptation project funded by development agencies and international funding bodies (the Green Climate Fund, for example). This would consist of an independent, transdisciplinary scientific monitoring of the project's implementation and results and assessment of the potential effectiveness of actions undertaken (including tracking), the risk of maladaptation, and the potential for residual risks and adaptation limits. The French Development Agency, for example, has started to explore this process by including a "Knowledge" component into its Adapt'Action Facility mechanism. In addition to providing feedback to funders and practitioners, this would represent a distinct opportunity to further engage local researchers, especially in the Global South, to better account for local realities at the crossroads of climate and development.

Second, a range of funding opportunities for adaptation research projects are already in place, both by national research agencies (*6*) and nonstate and nonpublic actors. For example, the AXA Research Fund, BNP-Paribas Foundation, and Rockefeller Foundation have become more active in identifying climate solutions (such as technological). These opportunities are critical to support research on the ground and new knowledge but, because they are project-oriented, are not well suited to support global-scale research structuring. One option is therefore to create a shared fund that is supplied by multiple organizations and specifically dedicated to the development of such international research structuring initiatives, starting with existing ones such as the Global Adaptation Mapping Initiative or the Gap-Track, just to name a few.

In addition to highlighting new scientific topics, the IPCC WGII warns that the solution space is shrinking with warming (*1*, *13*), so that a sharp acceleration in action is urgently needed. This raises multilevel challenges that the international community could help address, we argue, by setting up the enabling scientific conditions for the whole adaptation ecosystem (scientists, decision-makers, practitioners, and funders) to design and implement robust adaptation policy pathways. ■

REFERENCES AND NOTES

1. H.-O. Pörtner *et al*., Eds., "Summary for policy makers," *Climate Change 2022: Impacts, adaptation and vulnerability* (IPCC, 2022).
2. L. Berrang-Ford *et al*., *Nat. Clim. Chang*. **11**, 989 (2021).
3. UNEP, Adaptation Gap Report 2021 (UNEP, 2021).
4. B. C. O'Neill *et al*., "Key risks across sectors and regions," *Climate change 2022: Impacts, adaptation and vulnerability*, H.-O. Pörtner *et al*., Eds. (IPCC, 2022).
5. A. K. Magnan *et al*., *Nat. Clim. Chang*. **11**, 879 (2021).
6. R. H. Moss *et al*., *Science* **342**, 696 (2013).
7. V. Masson-Delmotte *et al*., Eds., *Special report on 1.5°C* (IPCC, 2018)
8. R. Pidcock, CarbonBrief 20 September 2016; https://www.carbonbrief.org/scientists-new-focus-1-5c-reshaping-climate-research/
9. E. Boyd *et al*., *One Earth* **4**, 1365 (2022).
10. T. R. Carter *et al*., *Glob. Environ. Change* **69**, 102307 (2021).
11. https://unfccc.int/sites/default/files/resource/AC_report_CMA_decision_GGA_0.pdf.
12. https://unfccc.int/sites/default/files/resource/ac2021_tp_gga.pdf.
13. M. Haasnoot, J. Lawrence, A. K. Magnan, *Science* **372**, 1287 (2021).

ACKNOWLEDGMENTS

This work was supported by the French Development Agency (grant GAP-Track_2021), "Programme d'investissements d'avenir" (ANR-10-LABX-14–01), and the Ademe (grant 20ESC0016) to A.K.M. and A.A.

10.1126/science.abq0737

POLICY FORUM

How trade policy can support the climate agenda

Ensure open markets for clean technologies and products

By **Michael Jakob, Stavros Afionis, Max Åhman, Angelo Antoci, Marlene Arens, Fernando Ascensão, Harro van Asselt, Nicolai Baumert, Simone Borghesi, Claire Brunel, Justin Caron, Aaron Cosbey, Susanne Droege, Alecia Evans, Gianluca Iannucci, Magnus Jiborn, Astrid Kander, Viktoras Kulionis, Arik Levinson, Jaime de Melo, Tom Moerenhout, Alessandro Monti, Maria Panezi, Philippe Quirion, Lutz Sager, Marco Sakai, Juan Sesmero, Mauro Sodini, Jean-Marc Solleder, Cleo Verkuijl, Valentin Vogl, Leonie Wenz, Sven Willner**

Economic analysis has produced ample insights on how international trade and climate policy interact (*1*). Trade presents both opportunities and obstacles, and invites the question of how domestic climate policies can be effective in a global economy integrated through international trade. Particularly problematic is the potential relocation of production to regions with low climate standards. Measures to level the playing field, such as border carbon adjustments (BCAs), may be justified for specific emissions-intensive and trade-exposed sectors but need to be well-targeted, carefully navigating tensions that can arise between the desire to respect global trade rules and the need to elaborate and implement effective national climate policies. The conformity of specific trade measures with international trade and climate change law is not entirely clear. Yet, clarity is needed to ensure that the industry actors affected will find the rules predictable and be able to adhere to them.

On the one hand, international trade can reduce emissions by providing incentives for countries with technological know-how to specialize in the production of clean goods and services, and countries with high renewable-energy potential to specialize in the clean production of energy-intensive goods. It can also accelerate the transfer and diffusion of clean technologies. Economic integration can enhance resilience to climate change by allowing regions affected by extreme weather to procure goods from unaffected trade partners (*2*), thus reducing price swings. It can also promote adaptation to slow-onset events by shifting production of affected sectors to regions experiencing lower climate-related productivity losses or even productivity gains.

On the other hand, current trade policies do not provide incentives for trade in clean products. For reasons unrelated to climate policy, they instead favor carbon-intensive imports. Raw materials and industries producing intermediate goods are the highest emitters of CO_2 per unit of value added. Yet, they face much lower tariffs and nontariff barriers (NTBs) compared to more complex products with lower carbon intensities, mainly because of counterlobbying by downstream industries. This bias has been shown to be equivalent to a negative carbon price of about US$90 per tonne of CO_2 on average (*3*)—the opposite of what a climate-supportive trade regime would require.

Differences in climate change mitigation costs across world regions could be decisive for the implementation of effective climate policies. Emissions resulting from the cross-border transport of goods are on average modest relative to emissions resulting from production. But carbon leakage through relocation of production from high- to low-regulation jurisdictions would be problematic. The mere anticipation of this type of carbon leakage could render ambitious emission targets politically difficult, if unmatched by other countries. Without clear and obvious benefits for the climate, governments will not be willing to sacrifice the competitiveness of key industries and jobs.

Yet the risk of trade-driven carbon leakage across the whole economy is generally found to be limited. Because of frictions to trade and because energy often constitutes a small share of overall costs, most industries would need very high carbon prices to justify relocation. Thus, leakage should not be used as a pretext to avoid ambitious climate policy or protect domestic industry from imports as it does not substantially compromise the possibility of unilaterally reducing global emissions (*4*).

CLUBS AND ADJUSTMENTS

The current debate on trade and climate change mainly revolves around border carbon adjustments. Recently, the European Commission laid out a proposal for a BCA that would impose the carbon costs borne by European Union (EU) producers on imports from countries with less stringent climate policies. The proposal is designed to levy the carbon price prevailing in the EU Emissions Trading System on imports of iron and steel products, fertilizers, aluminum, cement, and electricity (*5*). Export rebates, which could help domestic industries compete on a level playing field on the world market but which might violate international trade rules (*6*), are not part of the EU proposal. Canada and the United Kingdom are actively considering BCA regimes. Likewise, legislators in the United States have proposed border measures adjusting for differences in the costs of environmental regulation, not only carbon pricing.

The rules of the World Trade Organization (WTO) have received criticism for constraining the design of BCA schemes. Nevertheless, adhering to this set of rules facilitates conducting economic transactions across national boundaries. Countries will hence need to carefully navigate the tensions that can arise between the desire to respect global trade rules and the need to elaborate and implement effective national climate policies.

In addition to leveling the playing field, BCAs could constitute an important strategic instrument to deter free-riders who do not contribute to the global public good of climate change mitigation (*7*). Current efforts to reduce dependence on fossil fuel imports could increase political support for BCAs against exporters of fossil fuels and energy-intensive products. Yet, BCAs might put a substantial burden on other, especially low-income countries. Such inequitable outcomes would go against the United Nations Framework Convention on Climate Change (UNFCCC) principle of "common but differentiated responsibilities and respective capabilities." Political tensions and retaliation by trade partners would result in high economic costs and might also sour international relations to a degree that makes it harder to achieve progress in international climate cooperation. For instance, the European Commission expects that implementation of the proposed BCA will trigger a response by trade partners but remains agnostic as to the nature of this response.

BCAs could also be considered by a club of like-minded countries to confer favorable market access on members of the club, but trade barriers for those outside of it. The German G7 Presidency has vowed to advance the formation of a climate club of countries with comparable climate policies. Jointly adopted harmonized BCAs or

Author affiliations are listed in the supplementary materials.
Email: jakob@mcc-berlin.net

other trade restrictions could be a key element of a climate club (*8*). Harmonizing climate-related trade measures across countries with very different economic situations and climate policy ambitions, however, faces severe limitations. First, as WTO rules demand that foreign producers are not treated less favorably than domestic ones, the carbon price applied to imports cannot exceed the domestic carbon price level. As current carbon prices in, say, the EU and China differ by about a factor of 10, a common BCA in both jurisdictions would need to be set at no more than the Chinese level, which would be insufficient to protect EU producers from carbon leakage.

Second, US participation in a climate club is especially important, because the US constitutes the world's largest economy and second-largest emitter. However, adopting a carbon price in the US faces major political and legal obstacles. Owing to the legal provisions of the WTO, BCAs can only adjust for policies that put a price on carbon emissions, but not for the costs of other types of regulation, such as renewable portfolio standards. The possibility for a WTO-consistent BCA thus seems unlikely in the US. Third, very different approaches to carbon pricing in different countries—for instance, regarding covered sectors, the instrument used (carbon tax or emission trading), and whether domestic carbon prices are reimbursed to exporters—would further complicate cross-country harmonization of BCAs (*9*).

MAKING TRADE WORK FOR CLIMATE

BCAs are unlikely to provide sufficient incentives for joining a climate club, except for those few countries with substantial export shares in the sectors to be covered by a BCA. BCAs will thus have limited potential to force trade partners to adopt climate measures. In addition to the "stick" of a BCA, a climate club also needs "carrots" to make club membership attractive. A balanced mix of BCAs as enablers for ambitious national climate policies and positive incentives to ramp up climate action could help harmonize the trade and climate agendas.

A diplomatic BCA agenda

BCAs can ease opposition against climate policy from domestic industries at risk of carbon leakage. Any BCA should, however, be designed in a way that is least trade-restrictive to trade partners that are subject to this policy. The WTO rules generally favor open communication and multilateral efforts to reach negotiated solutions prior to the adoption of unilateral trade measures. Thus, international coordination helps align national climate policies with the WTO regime (*10*).

Trade policies adopted for environmental reasons can be compatible with the WTO's nondiscrimination principles under certain strict requirements. Even if a BCA violates the WTO's nondiscrimination rules, it can qualify for an exemption depending, in part, on whether countries engage in good faith efforts with their trade partners before adopting a BCA. Countries need to show that unilateral action was necessary or a last resort after having attempted to coordinate with trading partners. Consultations under the UNFCCC or WTO umbrella, as well as bilateral efforts, may satisfy this requirement. The WTO agreements also contain numerous transparency provisions with which a BCA regime needs to comply. This is particularly the case with respect to the purpose, the data used for calculations of the BCA, default values, exemptions and revenue use, and the review and phase-out criteria of the BCA. BCAs have so far not been tested before the WTO dispute settlement mechanism. Prior diplomatic efforts, even-handed application, and transparent administration could substantially increase the chances of a BCA regime to survive WTO scrutiny.

Use the WTO regime to reform and phase out fossil fuel subsidies

In 2020, global fossil fuel subsidies of US$351 billion constituted an important barrier for climate change mitigation (*11*). The WTO is in a good position to help promote fossil fuel subsidy reform. Given the trade effects of such subsidies, and following the example of ongoing talks on fisheries subsidies, the WTO's extensive rules on subsidies can help tackle fossil fuel subsidies, even though the WTO's dispute settlement system is currently not fully functional. Specifically, the WTO can strengthen transparency through improved notification by its members, counter-notification by other members, and by addressing fossil fuel subsidy reform in the Trade Policy Review Mechanism, a system under which WTO members peer-review their trade policies.

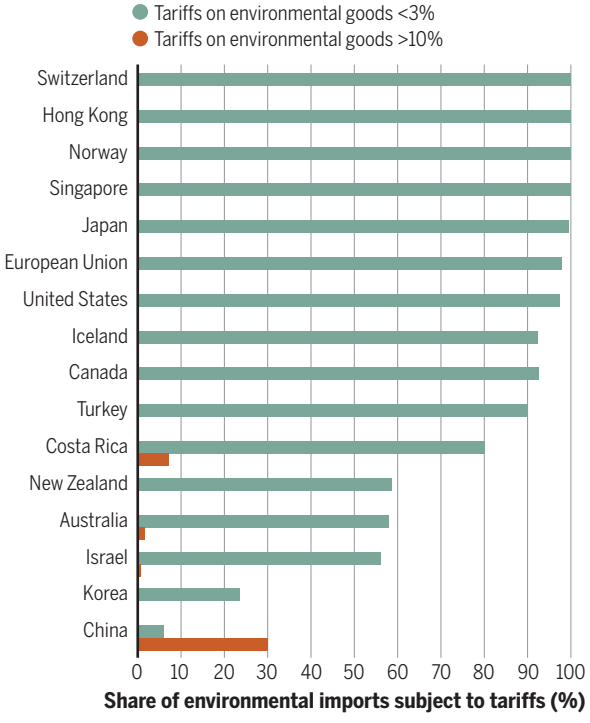
New rules—which could consist of an amendment of the Agreement on Subsidies and Countervailing Measures (ASCM) or a stand-alone agreement modeled after the proposed agreement on fisheries subsidies—could define a new category of prohibited subsidies. This may be limited to a subset of fossil fuel subsidies based on their trade and/or environmental effects and take into account the challenges faced by developing countries in subsidy reform (*12*). As changing (or agreeing to new) multilateral rules is challenging, progressive member states could push ahead with a plurilateral agreement, as some are doing in the context of the Agreement on Climate Change, Trade and Sustainability being negotiated by New Zealand, Costa Rica, Fiji, Iceland, Norway, and Switzerland.

Clear rules for settlement of disputes on renewable energy technologies

Lower trade barriers for climate-friendly goods decrease mitigation costs and could facilitate the achievement of more ambitious climate targets. As the international trade regime requires nondiscriminatory treatment of foreign products and producers, it also has the potential to accelerate the diffusion of clean energy technologies. A case in point is the successful legal challenges that have been raised against local content requirements in renewable energy support schemes. Such discriminatory tools require sourcing a certain share of value added from domestic producers, thereby favoring the lat-

Barriers to trade of clean technologies

Imports of environmental goods and services are subject to trade barriers in many nations. Some major economies apply tariffs of less than 3% (green bars) on almost all imports of environmental goods. For others, however, a substantial fraction of such imports is subject to tariffs of more than 3% and in some cases even more than 10% (orange bars) (*14*).



GRAPHIC: K. FRANKLIN/*SCIENCE*

ter over foreign firms and hampering global competition in the renewable energy sector, which may raise the costs of emission reductions. At the same time, such arrangements can also help establish competitive innovation capabilities and increase support for climate policies by domestic interest groups. Current geopolitical tensions and the concern about depending on imports of critical materials from authoritarian regimes could result in increased adoption of local content requirement schemes.

Temporary and adaptable local content requirements subject to regular assessment cycles might constitute a compromise between open markets for clean technologies and policy space for climate measures appropriate for the respective national context. To make the trade regime work for the climate, it will require a clear set of criteria to assess the compatibility of domestic support schemes for clean technologies with WTO principles.

Gaps can be identified in other areas that hinder the expansion of trade in renewable energy and green technologies. Examples include the lack of exceptions for climate-friendly subsidies under the ASCM, and the insufficiency of international trade rules to limit trade barriers for green technologies. Addressing such shortcomings through reform of the relevant WTO rules would be highly beneficial for trade in renewable energy and other climate-friendly technologies (*13*).

Expand efforts to achieve an agreement on environmental goods and services

The agenda to lower trade barriers for clean energy technologies should focus on tariff reductions for environmental goods, which are often still subject to high tariffs (see the figure) adopted to protect domestic industries from foreign competition. Environmental goods on average face higher trade barriers than less complex but more carbon-intensive products.

Negotiations on reducing tariffs on environmental goods have floundered for two decades, even though tariffs were already low among countries engaging in negotiations toward an Environmental Goods Agreement. Many are small "nuisance tariffs" (<3%) that could easily be removed to jump-start the process for such an agreement (*14*). This effort could be extended to higher tariffs and a longer list of goods that would include Environmentally Preferable Products. These include products that have been produced with lower environmental impacts than their conventional counterparts, such as organic food or sustainably grown timber products.

Packaging and labeling requirements or technical standards and norms can impose substantial NTBs for trade in environmental goods. NTBs related to labor market regulations, such as visa and work permit requirements, may hamper trade in environmental services—for instance, for the sustainable management of energy, water, and forest resources. Therefore, NTBs, which are difficult to detect and sometimes protectionist in intent, should be included in the agenda for mutual recognition.

Explore the potential of green materials clubs

Conditioning technology transfers to abatement efforts may strengthen incentives to join a climate club. A club of like-minded countries could go beyond imposing BCAs on nonmembers and also consider broader cooperation by implementing common green industrial policies, such as low-carbon requirements for climate-intensive globally traded basic materials, such as iron, steel, aluminum, cement, or fertilizers (*15*). The activities of a green materials club should start at relatively low ambition by developing long-term deep decarbonization visions or roadmaps for energy-intensive industries. In a second step, dedicated support for research, development, and diffusion of technologies and infrastructure planning can be applied in a harmonized manner by club members. More research is needed to gain a more comprehensive understanding of how technologies and policies can be adapted to specific country contexts.

The carrot for joining such a club, and thus adopting green industrial policies, would be getting access to future niche markets for green basic materials, in addition to avoiding (or at least reducing) the stick of BCAs. A green materials club could take on the form of a plurilateral agreement by a subset of early movers under the WTO. Coordinated efforts by a coalition of countries that are leaders in clean technologies would help establish such a green materials club. This would create a critical mass of participants to render the benefits of club membership attractive.

UNEXPLOITED POTENTIAL

Global trade has often been portrayed as an obstacle for climate policy, as it enables firms to move to export from regions with less ambitious environmental regulations. Such relocation of carbon-intensive production can be addressed by BCAs. Trade restrictions, however, can also bring along protectionist elements and only provide limited incentives for other countries to ramp up their climate ambition. BCAs need to be designed carefully according to clear and transparent rules in close dialogue with key trade partners to prevent political tensions that could imperil the entire climate agenda far beyond trade-related issues.

Trade policies supporting the climate agenda will need to pay particular attention to climate impacts on trade and the implications of climate-friendly trade for low-income countries. To support them, trade finance, investment, and aid for trade should be scaled up in areas related to climate change mitigation and adaptation, and additional research will be required to inform policy design.

Beyond unilateral policy approaches that include BCAs, the WTO regime offers a series of options to make the trade regime work for the climate. Unfortunately, up to now, this potential has not been exploited by WTO member states. The options discussed above will face substantial political obstacles. Yet growing awareness of the urgency to transform the global economy toward climate neutrality could give new impetus for restructuring the trade regime in a way that supports climate goals. ■

REFERENCES AND NOTES

1. J. R. Markusen, *J. Int. Econ.* **5**, 15 (1975).
2. M. Stevanović *et al.*, *Sci. Adv.* **2**, e1501452 (2016).
3. J. S. Shapiro, *Q. J. Econ.* **136**, 831 (2021).
4. J. Caron, "Empirical evidence and projections of carbon leakage: Some, but not too much, probably" in *Handbook on Trade Policy and Climate Change*, M. Jakob, Ed. (Edward Elgar, 2022).
5. European Commission, Proposal for a regulation of The European Parliament and of The Council establishing a carbon border adjustment mechanism (2021).
6. C. Böhringer, C. Fischer, K. E. Rosendahl, T. F. Rutherford, *Nat. Clim. Chang.* **12**, 22 (2022).
7. C. Coons, S. Peters, FAIR Transition and Competition Act of 2021: Protecting jobs, building resilience, and raising international climate ambition (2021).
8. W. Nordhaus, *Am. Econ. Rev.* **105**, 1339 (2015).
9. S. Tagliapietra, G. B. Wolff, *Nature* **591**, 526 (2021).
10. S. Droege, M. Panezi, "How to Design Border Carbon Adjustments" in *Handbook on Trade Policy and Climate Change*, M. Jakob, Ed. (Edward Elgar, 2022).
11. OECD & IEA, "Energy price surge underlines need to accelerate clean energy transitions rather than subsidise fossil fuels"; https://www.oecd.org/newsroom/energy-price-surge-underlines-need-to-accelerate-clean-energy-transitions-rather-than-subsidise-fossil-fuels.htm (2021).
12. H. van Asselt, T. Moerenhout, C. Verkuijl, "Using the Trade Regime to Phase out Fossil Fuel Subsidies" in *Handbook on Trade Policy and Climate Change*, M. Jakob, Ed. (Edward Elgar, 2022).
13. A. Monti, "International Trade Disputes on Renewable Energy" in *Handbook on Trade Policy and Climate Change*, M. Jakob, Ed. (Edward Elgar, 2022).
14. J. de Melo, J.-M. Solleder, "Towards An Environmental Goods Agreement Style (EGAST) agenda to improve the regime complex for Climate Change" *in Handbook on Trade Policy and Climate Change*, M. Jakob, Ed. (Edward Elgar, 2022).
15. M. Åhman, M. Arens, V. Vogl, "International cooperation for decarbonizing energy intensive industries – The case for a Green Materials Club" in *Handbook on Trade Policy and Climate Change*, M. Jakob, Ed. (Edward Elgar, 2022).

ACKNOWLEDGMENTS

We thank H. Colyer for language editing and A. Bosio for supporting the preparation of the manuscript.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo4207

10.1126/science.abo4207

REVIEW

Current global efforts are insufficient to limit warming to 1.5°C

H. Damon Matthews[1]* and Seth Wynes[1]

Human activities have caused global temperatures to increase by 1.25°C, and the current emissions trajectory suggests that we will exceed 1.5°C in less than 10 years. Though the growth rate of global carbon dioxide emissions has slowed and many countries have strengthened their emissions targets, current midcentury net zero goals are insufficient to limit global warming to 1.5°C above preindustrial temperatures. The primary barriers to the achievement of a 1.5°C-compatible pathway are not geophysical but rather reflect inertia in our political and technological systems. Both political and corporate leadership are needed to overcome this inertia, supported by increased societal recognition of the need for system-level and individual lifestyle changes. The available evidence does not yet indicate that the world has seriously committed to achieving the 1.5°C goal.

For the past two years governments and individuals worldwide have taken unprecedented action in response to the crisis created by the COVID-19 pandemic. Meanwhile global climate change has continued largely unabated and remains the leading global threat in the minds of both sustainability experts and business leaders (*1*). Despite calls for the prioritization of climate action measures as part of government pandemic recovery spending (*2*, *3*), a recent analysis concluded that only 6% of G20 stimulus spending has been directed toward areas that would lead to emissions reductions (*4*). Correspondingly, following a record drop in global CO_2 emissions between 2019 and 2020 in response to COVID lockdowns (*5*), emissions rebounded to close to 2019 levels in 2021—the largest single-year increase since the rebound following the 2008 to 2009 global financial crisis (*6*).

We review the current state of the climate system and global efforts to decrease CO_2 and other greenhouse gas emissions so as to meet the Paris Agreement's goal of limiting the global temperature increase from preindustrial levels to between 1.5°C and well below 2°C. We discuss both geophysical and socioeconomic barriers to achieving the more ambitious and less dangerous temperature goal of 1.5°C, as well as some of the more promising pathways toward achieving this outcome. Though there is evidence of mitigation progress in recent years and limiting warming to 1.5°C remains a possibility, we conclude that it is not a plausible outcome in response to the current level of global effort.

State of the climate system and the drivers of climate change

In its Sixth Assessment Report on the science of climate change, the Intergovernmental Panel on Climate Change (IPCC) reported that it is now "unequivocal that human influence has warmed the atmosphere, ocean, and land" (*7*). In this report the global temperature increase caused by human activities was estimated to be 1.07°C, representing the difference between the 2010 to 2019 average and the 1850 to 1900 baseline period (*8*). This human-caused (anthropogenic) warming was almost identical to the observed temperature increase of 1.09°C over the same period (*9*). As of June 2022 this anthropogenic warming has reached 1.25°C above the 1850 to 1900 baseline (*10*). Similarly, atmospheric CO_2 concentrations have increased steadily over the past six decades and now exceed 415 parts per million (ppm) (Fig. 1). The rate of observed temperature increase over the past 50 years—as well as the rates of increase of CO_2 concentrations and other greenhouse gases—are unprecedented in available records of past changes in climate (*9*).

Although observed increases in anthropogenic CO_2 and global temperature show no signs of slowing there is evidence that some key anthropogenic drivers of these observed changes have begun to diverge from the highest emission scenarios. After several decades of consistent increase, global CO_2 emissions from fossil fuels and land-use change have been approximately stable for the past decade (Fig. 2A). This trend in total CO_2 emissions is a combination of slightly increasing emissions from fossil fuel use and slightly decreasing (albeit highly uncertain) (*11*) emissions from land-use change (*6*). The year 2020 saw a 5.4% drop in fossil fuel emissions associated with COVID-19 lockdowns around the world (*5*, *6*), but this unprecedented decrease was followed by a substantial rebound of global emissions in 2021 back to 2019 levels. Notably, global CO_2 emissions trends over the past 5 years have diverged from the high-emission baseline scenarios that do not include climate policy efforts to curb emissions (Fig. 2A). This suggests that climate policies aimed at curbing CO_2 emissions have begun to have an effect, consistent with other analyses (*12*, *13*).

The trend in aggregated non-CO_2 forcing (the effect on climate caused by all other greenhouse gas and aerosol emissions) plotted in Fig. 2B shows a less clear tendency toward any particular future emissions scenario. There is some indication here that the highest emission scenario (SSP5-8.5) is the least consistent with observed forcing trends since 2015. However, it should be noted that the observed trend shown in Fig. 2B is dominated by the effect of decreased aerosol emissions (and consequent reduction of negative aerosol forcing) which explains most of the increase

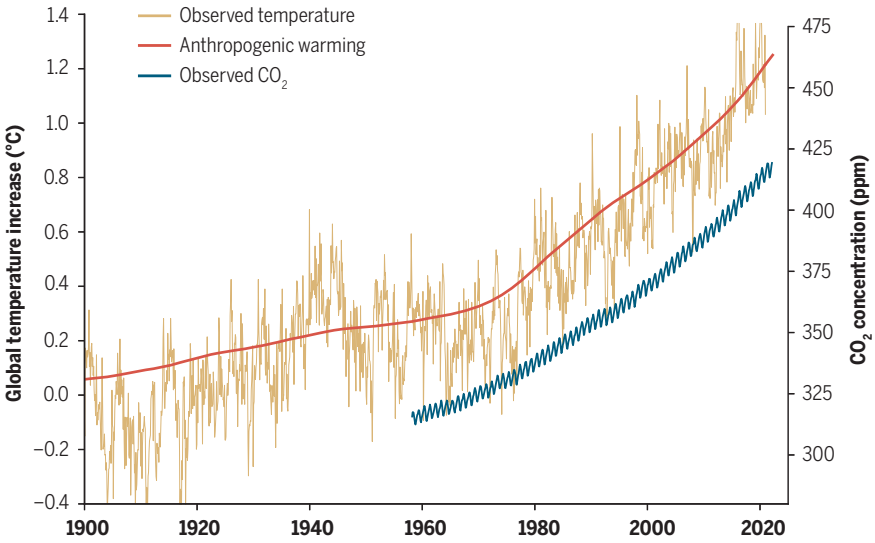


Fig. 1. Global temperature and atmospheric CO_2 change. Observed global temperature has increased by >1.2°C since the 1850 to 1900 baseline period (thin beige line; monthly temperature observations). Virtually all of this increase can be attributed to anthropogenic drivers (red line). The observed atmospheric CO_2 increase (blue line) is the primary driver of anthropogenic global warming.

[1]Department of Geography, Planning and Environment, Concordia University, Montreal, QB, Canada.
*Corresponding author. Email: damon.matthews@concordia.ca

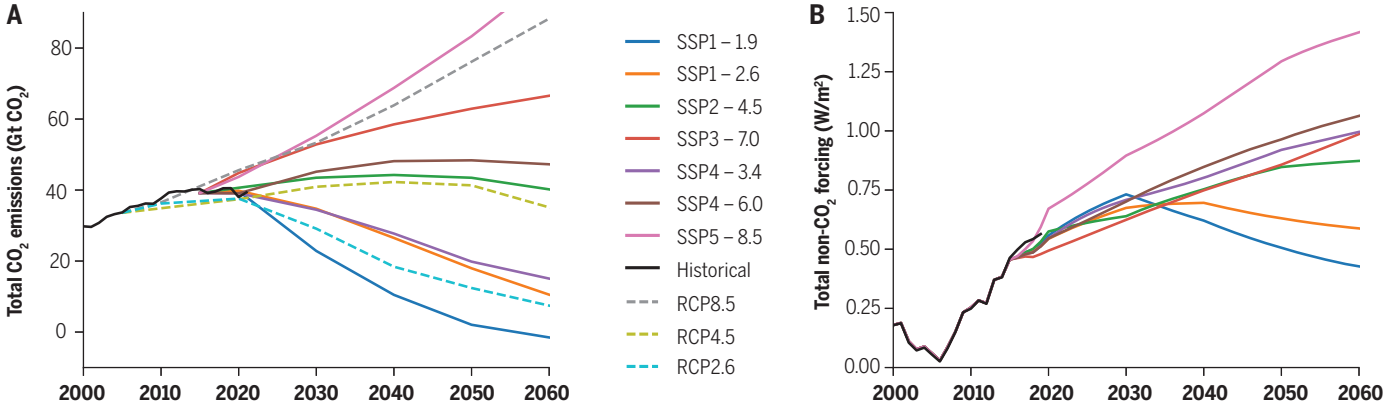


Fig. 2. Observed and projected CO_2 emissions and other anthropogenic climate drivers. (**A**) Global CO_2 emissions from fossil fuel combustion and land-use change have been approximately stable for the past decade and are no longer following projections of future emissions that do not include climate policy (SSP3-7.0, SSP5-8.5, and RCP8.5). (**B**) Aggregated non-CO_2 forcing (including other greenhouse gases and aerosols) has increased rapidly since 2005 largely in response to decreased aerosol forcing. Note that we have included the previous generation of RCP scenarios (in addition to the current SSP) in (A) for comparative purposes; non-CO_2 forcing estimates have been updated significantly in the past decade, however, so we have not included these older scenarios in (B).

in observed forcing since 2005 (*14*). Other individual forcing agents have not shown substantial departures from historical trends, though methane forcing has been trending upward in response to an increasing emission rate over the past decade (*15*).

Geophysical constraints on achieving 1.5°C

Despite indications that recent emissions have diverged from the highest emission scenarios, it is evident that emissions are not yet consistent with what would be required to meet the temperature goals of the Paris Agreement (1.5 to well below 2°C) (*16*). Extrapolation of the recent trends shown in Fig. 2 suggest approximately stable or slightly increasing greenhouse gas emissions over the coming decade, which is also consistent with analyses of the effect of current policies on emissions between now and 2030 (*17*–*20*). If current emission trends continue, global temperatures would be expected to continue to increase at a rate similar to the current warming rate (~0.24°C per decade; https://globalwarmingindex.org/), leading to temperatures that reach and cross the 1.5°C lower limit of the Paris temperature goal in about a decade (see https://climateclock.net/). This is consistent with the IPCC Sixth Assessment Report (AR6) which reported that the central estimate of the crossing time for 1.5°C in climate model simulations falls in the early 2030s for all future emission scenarios except SSP5-8.5 (*21*).

This 10-year timeline to 1.5°C is also consistent with recent estimates of the remaining carbon budget—the total quantity of future CO_2 emissions that would be consistent with a given amount of global temperature increase (*22*). The IPCC AR6 provided a median estimate of 500 billion metric tons (Gt) of CO_2 emissions from 2020 until the time of global net zero CO_2 emissions, which is equivalent to ~12.5 years of current annual CO_2 emissions (*15*). The authors also reported a 33rd to 67th percentile range of 400 to 650 Gt CO_2, corresponding to a 67 to 33% chance of not exceeding 1.5°C, respectively (*15*). This is comparable to the range of the remaining carbon budget reported in (*23*). However, this range does not include all relevant sources of uncertainty but only those related to the physical climate and carbon cycle processes that affect the climate response to CO_2 emissions. Other sources of geophysical uncertainties, such as those associated with non-CO_2 greenhouse gas and aerosol forcing, historical temperature estimates, and committed future warming collectively add a substantial additional spread to the reported remaining carbon budget values (*15*).

"The available evidence does not yet indicate that the world has seriously committed to achieving the 1.5°C goal."

Integrating these additional geophysical uncertainty sources into the remaining carbon budget distribution results in a somewhat smaller median estimate of 440 Gt CO_2 from 2020 onward but a much wider 33rd to 67th percentile range of 230 to 670 Gt CO_2 (*24*). Subtracting emissions for 2020 and 2021 from this median 1.5°C-remaining carbon budget leaves 360 Gt CO_2 for total emissions from 2022 onward; were global CO_2 emissions to decrease linearly to net zero from 2021 levels, respecting this remaining carbon budget would require net zero CO_2 emissions to be reached by the year 2040. This scenario of future CO_2 emissions would carry only a 50% chance of not exceeding 1.5°C, though it would provide much better odds of remaining below higher warming levels: a >80% chance of not exceeding 1.75°C and a >95% chance of remaining below 2°C (Fig. 3). This estimate of the remaining carbon budget is consistent with a typical 1.5°C-emission scenario that includes ambitious mitigation of both CO_2 and non-CO_2 emissions; however, the extent to which non-CO_2 emissions are successfully mitigated across scenarios adds an additional source of uncertainty that could either increase or decrease the estimate of the remaining carbon budget.

Misalignment of national targets with 1.5°C

The recent proliferation of net zero emissions targets among both nations and nonstate actors such as corporations and institutions (*25*) suggests a growing recognition of the need to eliminate net emissions in order to stabilize global temperatures (*26*). However, analyses of the alignment of current national targets with global temperature goals have unanimously concluded that targets are not yet sufficient to maintain temperatures within the 1.5 to well below 2°C goal of the Paris Agreement. In the Summary for Policymakers section of the AR6 climate mitigation report the IPCC stated that climate policies in place as of 2020 would not prevent emissions from continuing to rise and would lead to warming of 3.2°C by 2100 (*20*). Other projections of the effect of current climate policies have estimated end-of-century warming in the range of 2.5° to 3°C above 1850 to 1900 levels (*17*–*19*), which remains well outside of the Paris target window. Were all nations to meet their near-term (typically year 2030) targets, median end of century warming could be decreased to <2.5°C (*17*). If all nations were to additionally meet their long-term net zero targets, projected warming could be decreased

to 1.9°C—not yet reaching a "well below 2°C" scenario but one that could be characterized as "just below 2°C" (*27*).

Retaining the possibility of limiting warming to 1.5°C would require current net zero targets to be moved considerably closer in time, with a commensurate strengthening of near-term targets. This is particularly true for CO_2 emissions, which reach net zero in 1.5°C scenarios considerably earlier than other emissions (Fig. 4, A and C). Many nations have adopted a 2050 net zero target for greenhouse gas emissions (Fig. 4D) though most have not articulated an earlier target for net zero CO_2 emissions and in some cases have not provided sufficient clarity to determine whether the target applies to all greenhouse gases or only CO_2 emissions (*25*, *28*). If applied to CO_2 emissions, a 2050 net zero target would be roughly aligned with the typical net zero CO_2 emission year among 1.5°C scenarios, which generally occurs between 2046 and 2055 for the SR1.5 and AR6 scenarios (*20*, *29*). However, these 1.5°C scenarios assume substantial carbon dioxide removal in the second half of the century (*26*), with the result that the 1.5°C target is typically exceeded and then returned to at the end of the century. By contrast, a net zero CO_2 year of 2040 (consistent with the remaining carbon budget) would be expected to limit peak warming to 1.5°C with no subsequent assumptions regarding the use of carbon dioxide removal (Fig. 4A).

In the absence of national commitments to carbon removal, 2050 net zero emissions targets—whether applied to only CO_2 or to all greenhouse gases—are not sufficient to limit warming to 1.5°C. At best, the current set of national net zero targets (Fig. 4D) would lead to the 1.5°C target being exceeded without any expectation of returning below this temperature limit during the second half of the century. To avoid reliance on uncertain future carbon removal, nations should align their CO_2 emissions targets with a net zero year of 2040, which would be consistent with the median remaining carbon budget estimate for limiting peak warming to 1.5°C (*24*). Furthermore, unequal historical responsibility among nations (*30*, *31*) implies that wealthy nations should achieve net zero CO_2 emissions considerably earlier than the global average. Though the potential for a rapid transition to net zero varies among nations (*25*), a fair and ambitious 1.5°C-aligned net zero target for most wealthy industrialized nations would be one that sets out to achieve net zero CO_2 emissions earlier than the year 2040 (Fig. 4B).

Inertial barriers to achieving the 1.5°C goal

Uncertainty associated with the remaining carbon budget suggests that there is a non-negligible chance that we have already emitted enough CO_2 to cause temperatures to exceed 1.5°C. For example, Matthews *et al.* (2021) estimated a 17% chance that the 1.5°C remaining carbon budget was less than zero in 2020 (*24*). Such an outcome would be contingent on either very strong negative aerosol forcing (leading to substantial warming as aerosol emissions decrease) or a substantial amount of committed future warming associated with past CO_2 emissions. The best estimate of the so-called "zero-emissions commitment" (the additional warming that occurs following an abrupt cessation of CO_2 emissions) was close to zero in a recent climate model intercomparison (*32*). However, there was a large spread across models, indicating the possibility of additional warming that has not yet occurred in response to emissions that are already in the atmosphere. If climate inertia proves to be at the higher end of this uncertainty range, this would represent a strong physical constraint on the achievability of the 1.5°C target. However, the most likely estimates of the climate response to both past and future emissions leave a reasonable probability that 1.5°C of warming could still be avoided if both CO_2 and non-CO_2 emissions such as methane are aggressively mitigated over the coming decades.

Infrastructural or technological inertia likely constitutes a more influential constraint on the viability of the 1.5°C target. Power plants, industrial infrastructure, ships, aircraft, and land vehicles are all examples of expensive, fossil-fuel burning infrastructure with predictable lifespans that imply a certain quantity of future "committed emissions" (*33*). Committed emissions from existing energy infrastructure and proposed power plants in 2018 (846 Gt CO_2) already exceeded the remaining carbon budget for 1.5°C (*34*). More than half of committed emissions are derived from electricity generation and so achieving climate targets without extensive use of negative emissions technologies depends on some combination of carbon capture retrofits, canceling proposed power plants, or early retirement of existing power plants. For instance, in the case of coal power, maintaining a carbon budget consistent with 1.5°C would require retiring plants 10 to 30 years earlier than historical rates (*35*, *36*). Furthermore, continued construction of fossil fuel infrastructure results in greater committed emissions, which in turn would require more costly policies to still achieve temperature targets. For instance, manufacturing new internal combustion vehicles increases the likelihood that expensive policies will be required at a later point to incentivize early replacements (*37*, *38*).

Sectoral studies demonstrate that emissions commitments in other industries may not be as firm as in the power and transportation sectors. For instance, the magnitude of committed

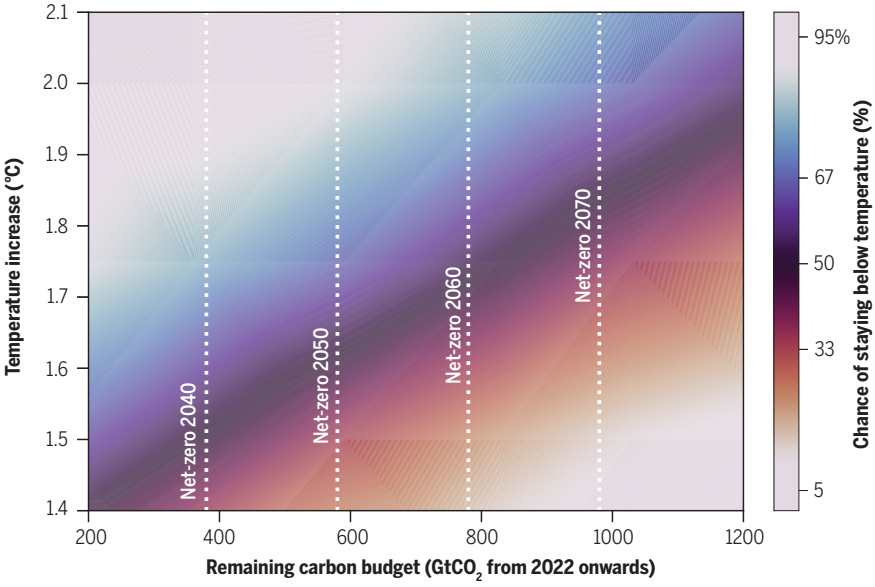


Fig. 3. Remaining carbon budget and the likelihood of temperature outcomes. A particular quantity of future cumulative CO_2 emissions can be associated with a likely range of temperature outcomes, reflecting the uncertainty in the remaining carbon budget associated with a particular temperature limit. For example, the median estimate of the 1.5°C remaining carbon budget [360 Gt CO_2 in this image, according to the distribution of Matthews *et al.* (2021) (*24*)], carries a 50% chance of remaining below 1.5°C but a >67% chance of remaining below 1.7°C and a >95% chance of not exceeding 2°C. This carbon budget distribution has a slightly smaller median estimate and considerably wider spread compared with those provided in the IPCC's SR1.5 and AR6 (*15*, *25*), owing to the inclusion of additional sources of geophysical uncertainty (*24*). Dashed vertical lines mark the cumulative CO_2 emissions and temperature outcomes that would be associated with a linear decline from 2021 CO_2 emission levels (~40 Gt CO_2 per year) to net zero CO_2 emissions at different years.

emissions from existing steel equipment (most recently estimated at 21 Gt CO_2e) is more driven by how quickly decarbonization technologies can be deployed than by the lifecycles of infrastructure (*39*). In the case of ships, and to a lesser extent, aircraft, the substitution of some low-carbon fuels might allow for decarbonization without the need for outright scrapping, though development of these technologies is expected to be slow, especially before 2030 (*40*, *41*).

Overall the literature on committed emissions from fossil fuel infrastructure does not suggest that achieving the 1.5°C target is physically impossible. Rather, this literature indicates the level of stringency of policies that would be required as well as the growing amount of technological infrastructure that would need to be retrofitted or retired earlier than its anticipated economic lifespan. It is therefore important to also consider a third source of inertia: current sociopolitical systems, which serve as an additional barrier to achieving the level of policy implementation that would be required for the 1.5°C target.

Sociopolitical inertia is evident in ongoing political and corporate support for fossil fuel exploration and extraction. To respect the 1.5°C remaining carbon budget most fossil fuel reserves cannot be extracted (*42*), implying the potential for enormous loss of economic value associated with current reserves and extraction infrastructure (*43*). This provides substantial incentives for fossil fuel–producing nations and companies to oppose decarbonization efforts (*44*). Other industries may also have incentives to oppose decarbonization efforts including electrical utilities, rail and steel companies, and the chemical industry (*45*, *46*).

In addition to the role of these corporate interests, public opinion is a key determinant of policy action in democracies (*47*). Despite growing concern and anxiety among some societal cohorts (*48*) there is evidence that positive changes in public opinion on climate change may be dependent on generational replacement (*49*). Furthermore, public opinion has not shifted consistently in response to increased natural disasters around the world (*50*–*52*). Public opinion is also subject to a status quo bias in which people tend to prefer things to stay as they are (*53*), which is undesirable in a situation in which rapid societal changes are necessary.

A similar status quo bias is seen in the decision making of political actors, and inertia is a well-known feature of institutions (*54*, *55*). Yet even when political actors and institutions are willing to make changes, crafting and passing effective legislation can take a meaningful share of a decade-long window. As an example, citizen climate assemblies are a democratic innovation that aim to engage citizens in decision-making related to climate policy (*56*). However, the successful French Convention Citoyenne pour le Climat took two years from announcement to passing of recommended legislation (*57*). It is therefore understandable that some have advocated for evaluating climate policies at least partially based on how quickly they can be implemented (*58*). Furthermore, even once legislation is passed implementation can be further delayed if national climate policies are opposed or obstructed by subnational actors (*59*).

Finally, changes in behavior including personal vehicle use and reduced meat consumption could result in large emissions cuts (*60*, *61*) although many climate-relevant behaviors are heavily tied to cultural practices. There is evidence that some social norms can change quickly (*62*) but habitual practices such as meat consumption and personal vehicle use are deeply entrenched (*63*–*65*).

"The major inertial forces that are likely to prevent this outcome are not climatic in nature, but rather political, technological, and social."

Physical limits to relevant technologies

The manufacturing of clean technologies, including electric vehicles, solar panels, and wind and geothermal energy systems, will require considerable resource expansion. Some resources such as copper for electrical transmission or steel for wind turbines are already produced at large scales whereas production of metals such as lithium, cobalt, and indium will need to expand rapidly to meet global needs (*66*). Despite these requirements there are no indications of strict geological limits that would prevent a clean energy transition (*67*, *68*). Instead, bottlenecks that could temporarily raise prices and slow decarbonization efforts are a more relevant concern (*69*). The ability to substitute materials and to recycle materials at end of life offers some margin for error, though less so in the short term (*70*).

The current regional concentration of critical mineral production and refining represents one source of potential disruption for the clean energy transition. Cobalt is a critical component of the lithium-ion batteries used in electric vehicles (*71*) but the world's largest producer of cobalt, the Democratic Republic of the Congo, is politically unstable in part because of the influence of international economic interests surrounding mineral extraction (*72*). Likewise, rare earth metals are necessary for a variety of renewable technologies and China is responsible for ~60 and ~90% of their production and refining, respectively (*73*). The lack of national or corporate diversity increases the risk of a supply chain failure. Other obstacles such as long lead times may also impede the rapid deployment of clean energy technologies. For example, it takes an average of 16 years for a mine to go from discovery to production and as such suppliers may struggle to match sudden escalations in demand (*73*).

Though widespread use of renewables will carry a substantial land footprint, land availability does not represent a hard technical constraint for Paris Agreement targets especially for nations with large emissions (*74*, *75*). Nevertheless, as renewable energy generation increases there will be fewer high-quality sites close to existing transmission infrastructure; this could in turn lead to land use competition in certain regions, even for energy-dense technologies such as solar (*76*). For technologies with low energy output per unit of land such as bioenergy, constraints are expected to be greater (*77*). Further, increased resource extraction and land competition are likely to drive sustainability challenges and human rights conflicts (*78*), suggesting the need for improved consultation with Indigenous peoples who have historically been harmed by energy projects (*79*, *80*).

The variable nature of wind and solar also poses issues for maintaining grid reliability during the clean energy transition. Fortunately, a variety of solutions exist including the expansion of electricity transmission infrastructure, overbuilding capacity, demand management, and incentivizing energy storage technologies (*81*). No energy storage technology can provide all grid services on its own but a combination of technologies (such as lithium-ion batteries, pumped hydroelectric storage, and reversible hydrogen fuel cells) can lend flexibility to renewable energy sources (*82*). Increased energy transmission infrastructure can act in a complementary way, balancing supply and demand across regions and thereby reducing the need for energy storage (*83*). More transmission can also allow for greater deployment of renewables in optimal locations (*82*), though new transmission projects often face public opposition (*84*).

Although a rapid clean energy transition will generate new obstacles, the difficulties are well understood compared with those posed by negative emissions technologies. Negative emissions technologies, which aim to remove carbon dioxide from the atmosphere and store it permanently, feature prominently in virtually all 1.5°C-compatible scenarios (*85*) (Fig. 4A). Limitations vary by the type of technology: bioenergy with carbon capture and storage (land, water, nutrients), direct air capture (energy), afforestation (land, permanence), soil carbon sequestration (permanence), and enhanced weathering (energy) (*86*, *87*). Given the early stages of development for these technologies it is unclear

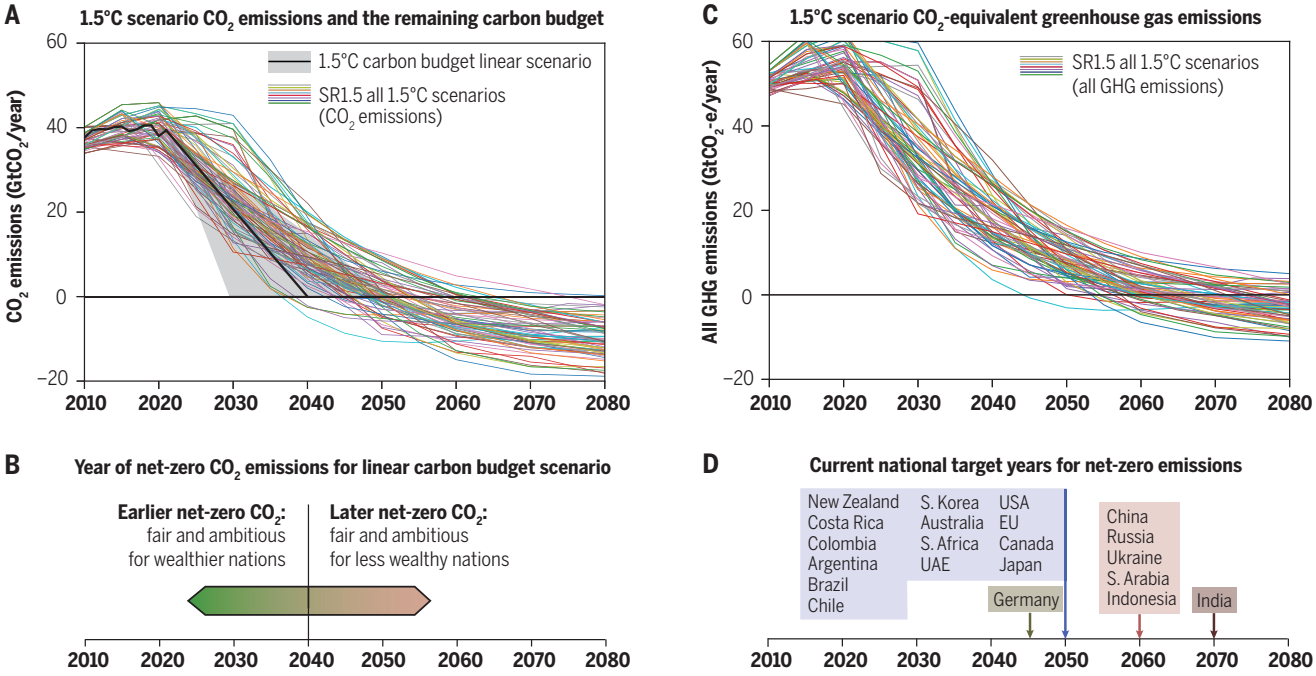


Fig. 4. Range of 1.5°C emission scenarios compared with the 1.5°C remaining carbon budget and current national net zero target years. (**A**) The median estimate of the remaining carbon budget from Matthews *et al*. (2021) would be consistent with a linear decline to net zero emissions by 2040 (solid black line; gray shading indicates the range of net zero years for the 33rd to 67th percentile of the remaining carbon budget). The set of 1.5°C scenarios from the IPCC 1.5°C Special Report (SR1.5) reach net zero CO_2 emissions around midcentury, generally followed by substantial amounts of net-negative emissions achieved through carbon dioxide removal. (**B**) To be aligned with the remaining carbon budget, net zero emissions pledges for CO_2 emissions should be centered on the year 2040, with wealthier nations (such as the US, EU, UK, Japan, and Canada) targeting earlier net zero CO_2 years in recognition of their higher historical responsibility. (**C**) Net zero years for all greenhouse gas emissions in SR1.5 scenarios occur later than that for CO_2 emissions only, and generally after the year 2050. (**D**) Current national net zero pledges in some cases apply to all greenhouse gas emissions, though some countries have used imprecise terms such as "carbon neutrality" or have not specified the scope of emissions coverage of their targets. In the absence of additional clarity and an explicit earlier net zero target for CO_2 emissions, these national targets are not well aligned with the remaining carbon budget. Rather they are generally aligned with the net zero CO_2 years of the SR1.5 scenarios, but without any subsequent commitments to net-negative emissions or reflection of the higher responsibility of wealthier nations.

how much carbon they will be able to store and how much society can rely on them to avoid a higher-temperature future (*88*).

Solutions in line with Paris targets

The IPCC AR6 report states with high confidence that achieving net zero CO_2 emissions will require counteracting the emissions from hard-to-abate sectors with carbon dioxide removal (*20*). However, carbon dioxide removal technologies are unproven at scale and as such there are dangers and moral issues associated with relying too heavily on these technologies to meet climate goals (*88*). Researchers have described several scenarios that are consistent with Paris Agreement goals but do not rely extensively on carbon dioxide removal (*89*–*91*). Features of these scenarios include the rapid decarbonization of energy grids, widespread electrification (e.g., heat pumps and electric vehicles), lifestyle changes (home energy conservation, diet shifts, and changes to travel patterns), efficiency improvements (e.g., home retrofits), reforestation, and the shuttering of unabated coal power plants (*89*–*91*).

In the very short term, policies such as car-free days in cities, lowered highway speed limits, and reduced business air travel could be implemented quickly with measurable effect (*92*). These should be paired with policies which take longer to develop and implement but must be pursued immediately, such as canceling fossil fuel subsidies. There is also still a need for research, development, and implementation of technologies in sectors in which emissions are more difficult to abate (*93*). In addition to governments, individuals also have a role to play in shifting their own behavior, demanding action from governments, and enacting changes in their workplaces (*94*, *95*). Likewise, corporations should lobby for action instead of fighting against it (*45*) while simultaneously moving beyond merely setting targets and disclosing emissions to assigning boardroom responsibility for climate action (*96*). Finally, nations themselves should take advantage of the ratcheting mechanism built into the Paris Agreement and strengthen their pledges in 2022 and beyond.

A rapid and just transition

In addition to the many benefits of keeping temperatures in line with Paris goals, climate solutions are often accompanied by substantial cobenefits such as reduced deaths attributable to air pollution (*97*) and unhealthy diets (*98*). However, rapid decarbonization is not without cost. For instance, upgrading internal combustion vehicles to heavier electric vehicles could result in more deadly collisions with pedestrians (*99*) and new jobs in renewable energy will not always be collocated with existing fossil fuel–centered employment (*100*). Likewise, a clean energy transition could lead to new patterns of exploitation (*78*) and recent failures in international climate financing indicate that the appetite for addressing similarly complex problems is limited (*101*). Ensuring a just transition will therefore require efforts at every level of government to improve the welfare of marginalized groups and others who might be negatively affected by energy system transformations.

Conclusions

Underlying all of these efforts is the emphasis on urgency. Major periods of societal mobilization in the past, including responses to world wars, may not be a direct analogy for positive climate action (*102*). However, these responses offer some useful policy parallels and clearly

establish the scale and speed required for rapidly decarbonizing societies (*103*). The societal response to the COVID-19 pandemic also clearly dwarfs the amount of effort and resources that have been directed toward climate mitigation in recent years. A global pandemic is an acute crisis whereas climate change is somewhat more chronic; the potential impacts of changes in climate on global societal stability and functioning are nevertheless equally if not more extreme if we do not succeed in limiting warming to manageable levels.

Scientific understanding of the climate response to emissions does not yet preclude the possibility of future emissions pathways that could limit either peak or long-term warming to 1.5°C. The major inertial forces that are likely to prevent this outcome are not climatic in nature, but rather political, technological, and social. Experts are pessimistic that the 1.5°C target will be met (*104*), and barring action that leads to a major upheaval in current norms, practices, and technologies, they will be correct. What is feasible or even possible to achieve is a function of knowledge, ability, desire, and effort. The current level of global effort is not sufficient to achieve what is required to limit warming to 1.5°C.

REFERENCES AND NOTES

1. Future Earth, "Global Risks Perceptions Report 2021," (Future Earth, Sustainability in the Digital Age, International Science Council, 2021); https://council.science/publications/global-risks-perceptions-report-2021/.
2. P. Le Billon, P. Lujala, D. Singh, V. Culbert, B. Kristoffersen, *Clim. Policy* **21**, 1347–1356 (2021).
3. M. Andrijevic, C.-F. Schleussner, M. J. Gidden, D. L. McCollum, J. Rogelj, *Science* **370**, 298–300 (2020).
4. J. M. Nahm, S. M. Miller, J. Urpelainen, *Nature* **603**, 28–31 (2022).
5. C. Le Quéré *et al*., *Nat. Clim. Chang.* **11**, 197–199 (2021).
6. P. Friedlingstein *et al*., *Earth System Science Data Discussions* **14**, 1917–2005 (2021).
7. V. Masson-Delmotte *et al*., in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021); pp. 3–32, (2021).
8. V. Eyring *et al*., in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021); pp. 423–552, (2021).
9. S. K. Gulev *et al*., in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021); pp. 287–422, (2021).
10. K. Haustein *et al*., *Sci. Rep.* **7**, 15417 (2017).
11. G. Grassi *et al*., *Nat. Clim. Chang.* **8**, 914–920 (2018).
12. C. Le Quéré *et al*., *Nat. Clim. Chang.* **9**, 213–217 (2019).
13. S. M. Eskander, S. Fankhauser, *Nat. Clim. Chang.* **10**, 750–756 (2020).
14. P. Forster *et al*., in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021); pp. 923–1054, (2021).
15. J. G. Canadell *et al*., in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021); pp. 673–8016, (2021).
16. M. J. Mace, Mitigation Commitments Under the Paris Agreement and the Way Forward. *Climate Law* **6**, 21–39 (2016).
17. Y. Ou *et al*., *Science* **374**, 693–695 (2021).
18. UNEP, "Emissions Gap Report 2021: The Heat Is On – A World of Climate Promises Not Yet Delivered," (2021).
19. Climate Action Tracker, The CAT Thermometer explained (2022); https://climateactiontracker.org/global/cat-thermometer/.
20. J. Skea *et al*., in *Climate Change 2022: Mitigation of Climate Change. Working Group III contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, (IPCC, 2022).
21. J.-Y. Lee *et al*., in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021); pp. 553–672, (2021).
22. H. D. Matthews *et al*., *Nat. Geosci.* **13**, 769–779 (2020).
23. J. Rogelj, P. M. Forster, E. Kriegler, C. J. Smith, R. Séférian, *Nature* **571**, 335–342 (2019).
24. H. Damon Matthews *et al*., An integrated approach to quantifying uncertainties in the remaining carbon budget. *Communications Earth & Environment* **2**, 7 (2021).
25. J. Rogelj, O. Geden, A. Cowie, A. Reisinger, *Nature* **591**, 365–368 (2021).
26. H. D. Matthews, K. Caldeira, Stabilizing climate requires near-zero emissions. *Geophys. Res. Lett.* **35**, L04705 (2008).
27. M. Meinshausen *et al*., *Nature* **604**, 304–309 (2022).
28. Climate Action Tracker, <https://climateactiontracker.org/global/cat-net-zero-target-evaluations/ (2021).
29. J. Rogelj *et al*., in *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, V. Masson-Delmotte *et al*., Eds. (2018).
30. H. D. Matthews, *Nat. Clim. Chang.* **6**, 60–64 (2016).
31. P. Baer, T. Athanasiou, S. Kartha, E. Kemp-Benedict, in *Environmental Rights*. (Routledge, 2017), pp. 75-89.
32. A. H. MacDougall *et al*., *Biogeosciences* **17**, 2987–3016 (2020).
33. S. J. Davis, K. Caldeira, H. D. Matthews, *Science* **329**, 1330–1333 (2010).
34. D. Tong *et al*., *Nature* **572**, 373–377 (2019).
35. R. Fofrich *et al*., *Environ. Res. Lett.* **15**, 094064 (2020).
36. R. Y. Cui *et al*., *Nat. Commun.* **10**, 4759 (2019).
37. A. F. Alarfaj, W. M. Griffin, C. Samaras, *Environmental Research Letters* **15**, 0940c0942 (2020).
38. D. R. Keith, S. Houston, S. Naumov, *Environ. Res. Lett.* **14**, 021001 (2019).
39. V. Vogl, O. Olsson, B. Nykvist, *Joule* **5**, 2646–2662 (2021).
40. S. Gössling, C. Lyle, *Transp. Rev.* **41**, 643–658 (2021).
41. S. Bullock, J. Mason, J. Broderick, A. Larkin, *BMC Energy* **2**, 1–16 (2020).
42. D. Welsby, J. Price, S. Pye, P. Ekins, *Nature* **597**, 230–234 (2021).
43. J. F. Mercure *et al*., *Nat. Clim. Chang.* **8**, 588–593 (2018).
44. T. A. Hansen, *Renew. Sustain. Energy Rev.* **158**, 112144 (2022).
45. R. J. Brulle, *Sociol. Inq.* **91**, 603–624 (2021).
46. T. Culhane, G. Hall, J. T. Roberts, *Energy Res. Soc. Sci.* **79**, 102114 (2021).
47. F. R. Baumgartner, B. D. Jones, *Agendas and Instability in American Politics* (Univ. of Chicago Press, 1993).
48. J. Wu, G. Snell, H. Samji, *Lancet Planet. Health* **4**, e435–e436 (2020).
49. M. Motta, *Int. J. Public Opin. Res.* **33**, 477–489 (2020).
50. C. Hazlett, M. Mildenberger, *Am. Polit. Sci. Rev.* **114**, 1359–1365 (2020).
51. C. Zanocco, H. Boudet, R. Nilson, J. Flora, *Glob. Environ. Change* **59**, 101984 (2019).
52. C. A. Ogunbode, C. Demski, S. B. Capstick, R. G. Sposato, *Glob. Environ. Change* **54**, 31–39 (2019).
53. C. Lang, M. Weir, S. Pearson-Merkowitz, *Environ. Res. Lett.* **16**, 054076 (2021).
54. R. J. Brulle, K. M. Norgaard, *Env. Polit.* **28**, 886–908 (2019).
55. C. J. Uittenbroek, *J. Environ. Policy Plann.* **18**, 161–176 (2016).
56. R. Willis, N. Curato, G. Smith, *Wiley Interdiscip. Rev. Clim. Change* **13**, 759 (2022).
57. C. Cherry *et al*., Citizens' climate assemblies: Understanding public deliberation for climate policy. (2021).
58. L. H. Goulder, *Rev. Environ. Econ. Policy* **14**, 143–156 (2020).
59. K. Jörgensen, A. Jogesh, A. Mishra, *J. Integr. Environ. Sci.* **12**, 235–245 (2015).
60. L. Cozzi, A. Petropoulos, "Growing preference for SUVs challenges emissions reductions in passenger car market," (International Energy Agency, 2019).
61. M. N. Hayek, H. Harwatt, W. J. Ripple, N. D. Mueller, *Nat. Sustain.* **4**, 21–24 (2021).
62. D. Centola, J. Becker, D. Brackbill, A. Baronchelli, *Science* **360**, 1116–1119 (2018).
63. K.-C. Nam, C. Jo, M. Lee, *Meat Sci.* **86**, 95–102 (2010).
64. G. Mattioli, C. Roberts, J. K. Steinberger, A. Brown, *Energy Res. Soc. Sci.* **66**, 101486 (2020).
65. J. Graça, C. A. Godinho, M. Truninger, *Trends Food Sci. Technol.* **91**, 380–390 (2019).
66. K. Hund, D. La Porta, T. P. Fabregas, T. Laing, J. Drexhage, Minerals for climate action: the mineral intensity of the clean energy transition. *World Bank* (2020).
67. A. Månberger, B. Stenqvist, *Energy Policy* **119**, 226–241 (2018).
68. J. Lee *et al*., *Renew. Sustain. Energy Rev.* **124**, 109789 (2020).
69. A. Valero, A. Valero, G. Calvo, A. Ortego, *Renew. Sustain. Energy Rev.* **93**, 178–200 (2018).
70. T. Junne, N. Wulff, C. Breyer, T. Naegler, *Energy* **211**, 118532 (2020).
71. X. Fu *et al*., *Environ. Sci. Technol.* **54**, 2985–2993 (2020).
72. M. B. Rapanyane, *Contemp. Soc. Sci.* **17**, 117–128 (2021).
73. IEA, "The Role of Critical Minerals in Clean Energy Transitions," in *World Energy Outlook Special Report* (2021).
74. S. Baruch-Mordo, J. Kiesecker, C. Kennedy, J. Oakleaf, J. Opperman, *Environ. Res. Lett.* **14**, 024013 (2019).
75. E. Larson *et al*., Net-Zero America (Princeton Univ., 2021).
76. D.-J. van de Ven *et al*., *Sci. Rep.* **11**, 2907 (2021).
77. W. V. Reid, M. K. Ali, C. B. Field, *Glob. Change Biol.* **26**, 274–286 (2020).
78. B. K. Sovacool *et al*., *Science* **367**, 30–33 (2020).
79. E. Zárate-Toledo, R. Patiño, J. Fraga, *Energy Res. Soc. Sci.* **54**, 1–11 (2019).
80. D. Avcı, C. Fernández-Salvador, *Extr. Ind. Soc.* **3**, 912–921 (2016).
81. M. R. Shaner, S. J. Davis, N. S. Lewis, K. Caldeira, *Energy Environ. Sci.* **11**, 914–925 (2018).
82. L. Clarke *et al*., in *IPCC, 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. P.R. Shukla *et al*., Eds. (Cambridge Univ. Press, 2022).
83. D. P. Schlachtberger, T. Brown, S. Schramm, M. Greiner, *Energy* **134**, 469–481 (2017).
84. G. Hoberg, *The Resistance Dilemma: Place-based Movements and the Climate Crisis*. (MIT Press, 2021).
85. S. Low, S. Schäfer, *Energy Res. Soc. Sci.* **60**, 101326 (2020).
86. J. C. Minx *et al*., *Environ. Res. Lett.* **13**, 063001 (2018).
87. P. Smith *et al*., *Nat. Clim. Chang.* **6**, 42–50 (2016).
88. K. Anderson, G. Peters, *Science* **354**, 182–183 (2016).
89. D. P. van Vuuren *et al*., *Nat. Clim. Chang.* **8**, 391–397 (2018).
90. IEA, Net Zero by 2050: A Roadmap for the Global Energy Sector. (2021); https://www.iea.org/reports/net-zero-by-2050
91. A. Grubler *et al*., *Nat. Energy* **3**, 515–527 (2018).
92. IEA, "A 10-Point Plan to Cut Oil Use," (International Energy Agency, 2022).
93. P. Fennell, J. Driver, C. Bataille, S. J. Davis, *Nature* **603**, 574–577 (2022).
94. K. S. Nielsen, K. A. Nicholas, F. Creutzig, T. Dietz, P. C. Stern, *Nat. Energy* **6**, 1011–1016 (2021).
95. S. Wynes, M. Motta, S. D. Donner, *One Earth* **4**, 363–371 (2021).
96. S. Dietz *et al*., *Nat. Clim. Chang.* **8**, 1072–1075 (2018).
97. K. Vohra *et al*., *Environ. Res.* **195**, 110754 (2021).
98. W. Willett *et al*., *Lancet* **393**, 447–492 (2019).
99. B. Shaffer, M. Auffhammer, C. Samaras, *Nature* **598**, 254–256 (2021).
100. S. Pai, H. Zerriffi, J. Jewell, J. Pathak, *Environ. Res. Lett.* **15**, 034065 (2020).
101. J. T. Roberts *et al*., *Nat. Clim. Chang.* **11**, 180–182 (2021).
102. J. Kester, B. K. Sovacool, *Energy Policy* **104**, 50–55 (2017).
103. L. L. Delina, M. Diesendorf, *Energy Policy* **58**, 371–380 (2013).
104. J. Tollefson, *Nature* **599**, 22–24 (2021).

ACKNOWLEDGMENTS

Funding: H.D.M. acknowledges funding from the Natural Sciences and Engineering Research Council of Canada. The contributions by S.W. were supported by a fellowship from the Social Sciences and Humanities Research Council of Canada. **Author contributions:** H.D.M. led the study and wrote the sections related to the state of the climate system, geophysical constraints, and the alignment of targets with global temperature goals. S.W. led the writing of sections in the second half of the paper and provided input on other sections. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data displayed in this paper is taken from previously published studies or publicly available data repositories as noted in the manuscript text.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

Submitted 4 April 2022; accepted 17 May 2022
10.1126/science.abo3378

REVIEW

Harnessing the potential of nature-based solutions for mitigating and adapting to climate change

Nathalie Seddon

Although many governments, financial institutions, and corporations are embracing nature-based solutions as part of their sustainability and net-zero carbon strategies, some nations, Indigenous peoples, local community groups, and grassroots organizations have rejected this term. This pushback is fueled by (i) critical uncertainties about when, where, how, and for whom nature-based solutions are effective and (ii) controversies surrounding their misuse in greenwashing, violations of human rights, and threats to biodiversity. To clarify how the scientific community can help address these issues, I provide an overview of recent research on the benefits and limits of nature-based solutions, including how they compare with technological approaches, and highlight critical areas for future research.

Nature-based solutions are actions that involve people working with nature, as part of nature, to address societal challenges, providing benefits for both human well-being and biodiversity (*1*, *2*). Climate change is one such key societal challenge. Examples of nature-based solutions that can help people adapt to climate change impacts include community-led protection, restoration, and management of natural and seminatural ecosystems within river catchments or along coastlines to protect against flooding and erosion; sustainable management of working lands to sustain or enhance crop yields during unpredictable growing conditions; and creation of forests, parks, and wetlands within and around urban areas for cooling and to reduce flood risk. In addition to supporting human adaptation, such actions can increase carbon storage to help mitigate climate change and protect biodiversity (*3*) (Fig. 1).

The rapid rise in the prominence of nature-based solutions on policy, research, and business agendas (Fig. 2) (*2*) is based on their perceived or documented effectiveness, readiness, scalability, and affordability relative to those of technological solutions (*3*). It also reflects broader recognition of the interdependency of societal well-being and ecosystem health (*4*). Yet, in the final hours of the negotiations at the 26th Conference of the Parties (COP26) to the United Nations Framework Convention on Climate Change, the term "nature-based solution" was removed from the decision text (i.e., the Glasgow Climate Pact). What remained was a commitment by the pact's 197 signatories to recognize "the critical role of protecting, conserving and restoring nature and ecosystems in delivering benefits for climate adaptation and mitigation, while ensuring social and environmental safeguards" (*5*). Although it could be claimed that nature-based solutions were included in all but name, COP26 revealed that whereas many organizations and governments are embracing the approach as an essential tool for tackling climate change, others, particularly grassroots organizations, have dismissed it as a dangerous distraction from systemic change (*6*). Contrasting narratives partly reflect uncertainty in the underlying evidence as well as controversies on how nature-based solutions are conceptualized and implemented. Here, I review recent research on the effectiveness of nature-based solutions to climate change mitigation and adaptation and provide an overview of how an improved evidence base, supported by interdisciplinary research and traditional knowledge, can and must help address these issues.

"Nature-based solutions can make an important contribution to reaching net-zero...but only if combined with other climate solutions..."

The role of nature-based solutions in addressing climate change

There is general scientific consensus that to achieve the Paris Agreement ambition of limiting mean temperature increase to 1.5°C, the global economy will require substantial, rapid reductions in greenhouse gas (GHG) emissions, reaching net-zero CO_2 by around 2050 and likely needing ongoing net-CO_2 removals thereafter (*7*). Nature-based solutions play a key role by reducing the release of GHGs from the agriculture, forest, and land-use sector—which currently account for ~22% of annual GHG emissions—and by protecting and enhancing the carbon sink on land and in the sea (*7*). Specifically, improving the management of working lands and seas (e.g., croplands, pastures, timberlands, and aquatic systems) reduces emissions of CO_2, methane, and nitrous oxide and sequesters carbon; restoring native vegetation across rural landscapes and planting trees in cities enhances CO_2 removal from the atmosphere; and effectively protecting intact forests, grasslands, and wetlands keeps critical reserves of carbon locked in the biosphere (*7*).

Estimates of the extent to which nature-based solutions can cool the planet vary considerably. To date, there have been more than 30 published estimates for the global contribution of nature-based solutions to climate change mitigation (*8*). These estimates include terrestrial and coastal ecosystems but exclude oceans, for which knowledge of carbon fluxes is limited (*9*). The large range in estimates of land-based mitigation globally [from ~100 to more than 800 billion tonnes (Gt) of accumulated CO_2 by the end of the century] reflects differences in studies' methods and assumptions (*8*). High estimates assume that society is willing to pay a high price for carbon (>US$100 per tonne), consider few biophysical or socioeconomic constraints on implementation, include a wide range of ecosystems, and not distinguish between afforestation (i.e., planting forests where they did not previously exist) and reforestation approaches (*8*). Conversely, low estimates assume a carbon price <US$100 per tonne, consider stricter biophysical and socioeconomic limits on where nature-based solutions can be implemented, and/or include a smaller subset of ecosystems or types of interventions (*8*). Including net climate effects in models also increases the uncertainty of global estimates of climate change mitigation potential of nature-based solutions (*8*, *10*). Although nature-based solutions generally increase carbon storage, they can also have local biophysical or biochemical effects that increase temperatures (for example, by decreasing albedo), depending on latitude and type of vegetation.

Even the most constrained estimates of the contribution of land-based nature-based solutions to global climate change mitigation are highly uncertain. These estimates do not consider the risk of impermanence, as climate change and other anthropogenic stressors can undermine ecosystem health (Fig. 1) (*11*). Nor do they account for the serious problem that scaling up of nature-based solutions in one region can result in the export of ecosystem loss and damage to another (a phenomenon termed "leakage"). Leakage is especially problematic when it results in biodiversity loss through the degradation and destruction of native vegetation elsewhere (*12*). In view of such issues, a conservative potential for nature-based solutions on land globally to contribute to climate change mitigation is around 100 to 200 Gt of CO_2 by 2100 (*8*) or, at most, 11.5 Gt of CO_2 equivalents per year up to 2050 (a CO_2 equivalent is the number of tonnes of CO_2 emissions with the same global warming potential as 1 tonne of another greenhouse gas) (*7*). This corresponds to a ~0.3°C reduction

Nature-based Solutions Initiative, Department of Biology, University of Oxford, Oxford OX1 3SAZ, UK.
Corresponding author. Email: nathalie.seddon@zoo.ox.ac.uk

in mean global temperature if temperature rise since preindustrial times peaks at 2°C toward the end of this century (*13*). In other words, though it is important, the contribution of nature-based solutions to global cooling is much smaller than what must also be achieved through drastic cuts in the use of fossil fuels (*14*).

In recognition of the fact that substantial climate change has already occurred and further warming is inevitable, Paris Agreement signatories have committed to "enhancing adaptive capacity, strengthening resilience and reducing vulnerability to climate change" (*15*). Although no single metric can track adaptation progress, and it is as much a process as an outcome, there are many examples of successful adaptation at the local level (Fig. 3). However, a recent systematic review of 1682 publications found that humans are failing to implement adaptation strategies, other than by making minor adjustments such as developing drought-resistant crops (*16*). Consequently, there is a large and widening gap between climate change adaptation needs and action. This is especially true in low-income countries, where there is a lack of international climate finance and adaptation costs are 5 to 10 times as high as the current flows of public funding (*17*).

Nature-based solutions can help bridge this gap. People have been working with nature to adapt to environmental changes since time immemorial. Nature-based solutions harness this process and contribute to different aspects of adaptation (*3*, *18*, *19*). In particular, there is robust evidence to show that nature-based solutions can reduce direct exposure to climate change impacts. For example, restoring native ecosystems can promote healthy soil and vegetation that reduce the risks of floods, droughts, and landslides by increasing infiltration and storage of water, stabilizing slopes and shores, and attenuating wave energy (*20*–*22*) (Fig. 1). Introducing green and blue infrastructure (such as bioswales, green roofs, and constructed wetlands) into urban areas can help moderate heatwaves and reduce flooding (*22*, *23*). There is also evidence that nature-based solutions can reduce sensitivity—that is, the degree to which people are affected by climate impacts. For example, nature-based agricultural practices such as agroforestry can sustain or enhance yields in drier or more unpredictable climates (*24*, *25*). Further, implementing nature-based solutions can build capacity to innovate or adjust in response to changing conditions, thereby supporting adaptation as a process (*19*).

Nature-based solutions can complement engineered approaches to address climate mitigation and adaptation, with some key advantages. Technological approaches for GHG removal, such as direct air capture and enhanced weathering, are effective but are currently expensive and energy intensive (*26*). Similarly, structures designed to reduce climate change impacts, such as seawalls and dikes, can be constructed quickly and are efficient in mitigating specific hazards. However, construction brings social and environmental damages locally, and the financial costs of maintaining or upgrading engineered structures can be very high [e.g., US$12 billion to US$71 billion per year to protect against sea level rise with dikes globally (*27*)]. By contrast, nature-based solutions can take time to establish, and their efficacy can vary with the intensity and frequency of threats and the resilience of ecosystems to climate impacts. Yet as living, evolving systems, nature-based solutions have the potential for self-repair and adaptation and may keep pace with a changing climate—for example, some studies suggest that oyster reefs and mangroves can track sea level rise (*3*, *28*). Thus, although initial implementation costs of nature-based solutions may be high (*29*), maintenance costs over the long term may be lower than the cost of engineered alternatives.

Few studies directly contrast the efficacy of nature-based and other approaches, but some have shown that, when their long-term and broader benefits are taken into account, nature-based and/or hybrid solutions can have benefit/cost ratios that are many times as large as those

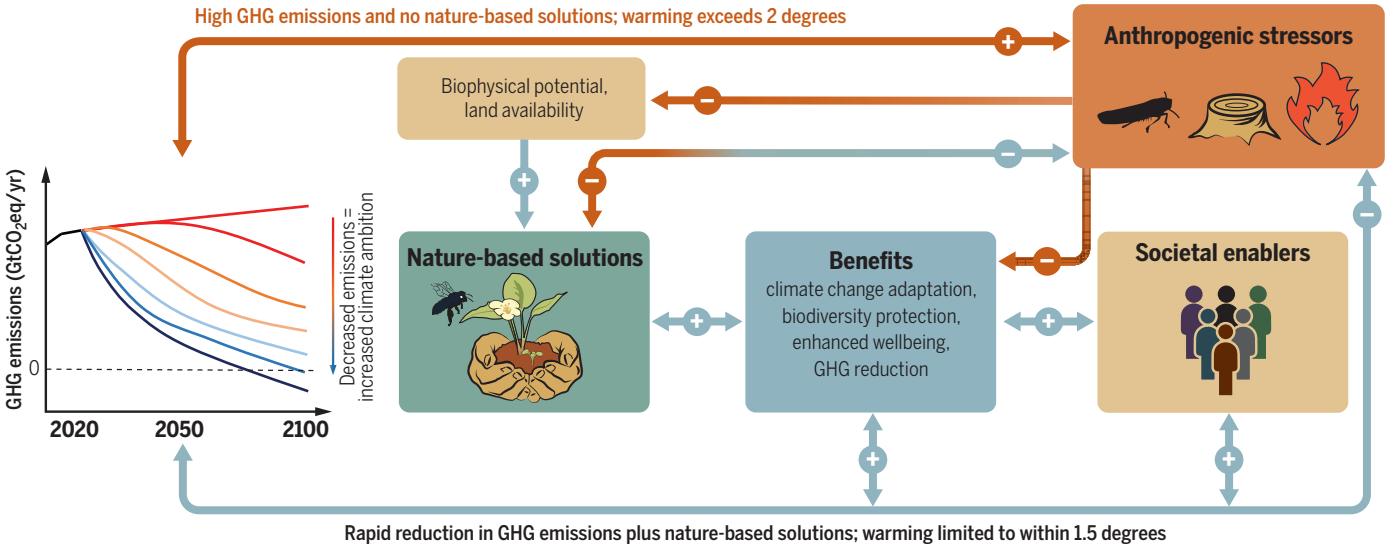

High GHG emissions and no nature-based solutions; warming exceeds 2 degrees
Anthropogenic stressors
Biophysical potential, land availability
GHG emissions (GtCO2eq/yr)
0
2020
2050
2100
Decreased emissions = increased climate ambition
Nature-based solutions
Benefits
climate change adaptation, biodiversity protection, enhanced wellbeing, GHG reduction
Societal enablers
Rapid reduction in GHG emissions plus nature-based solutions; warming limited to within 1.5 degrees

Fig. 1. Nature-based solutions and the key factors that influence their effectiveness. Nature-based solutions are "place-based partnerships between people and nature" that are underpinned by biodiversity and implemented by and for local communities (*2*). When well designed and carefully implemented, nature-based solutions bring multiple benefits (blue box), although the extent to which they can be deployed depends on biophysical potential (including net climate effects) and land availability, as well as a range of societal enablers (tan box on the right) that influence the implementation and upkeep of nature-based solutions over time. These enablers include good governance, secure land tenure, sustainable livelihoods and finance, and nature-positive values. However, nature-based solutions must be implemented in tandem with rapid reductions in GHG emissions (high climate policy ambition, represented by the blue and purple lines in the lower part of the graph on the left). Otherwise, the warming that results from high and increasing GHG emissions (low climate policy ambition, represented by the orange and red lines in the upper part of the graph on the left) will turn the biosphere into a net source of GHGs through the increased frequency and/or intensity of fires, extreme events, and pests (orange box), although these impacts can be reduced through careful adaptive management of nature-based solutions. The GHG emissions trajectories shown here are derived from scenarios published in the latest report from Working Group III of the Intergovernmental Panel on Climate Change (*7*). Orange arrows denote harmful effects, whereas pale blue arrows denote beneficial effects; + and − symbols indicate where effects are achieved by increasing (+) or decreasing (−) the target factors.

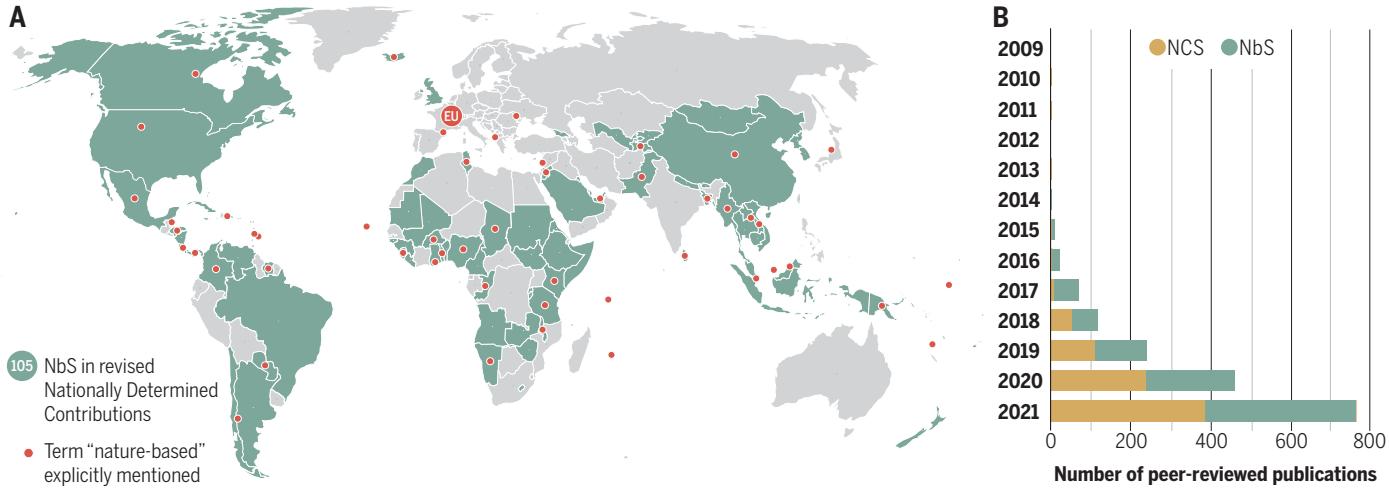


Fig. 2. Global traction of nature-based solutions. Recent years have seen a rapid rise in the prominence of nature-based solutions on policy (**A**), research (**B**), and business agendas (*2*). There has been a steep increase in the number of peer-reviewed articles identified by Google Scholar that use the term "nature-based solution" (NbS) or the related term "natural climate solution" (NCS) (which refers specifically to nature-based solutions for climate change mitigation) in their titles, abstracts, or keywords (B). Meanwhile, more than 80% of revised Paris Agreement climate pledges [nationally determined contributions (NDCs)] include nature-based solutions in their mitigation and/or adaptation plans (A), with more than 40% explicitly using the term "nature-based" (red dots). The large red dot for EU indicates that the NDC was submitted by the European Union as a whole (in contrast to countries that sent individual NDCs). Note that some nations (such as Peru) include nature-based solutions in their national adaptation plans but not their NDCs, whereas others (such as Bolivia, Côte d'Ivoire, and Madagascar) included nature-based solutions in their first NDC but had not submitted an updated version by the end of 2021 (and hence are gray in this figure). For more information, visit www.nbspolicyplatform.org.

of engineered alternatives (*3*, *30*). Further, there are indications that nature-based solutions are more cost effective than engineered approaches in lower-intensity hazard scenarios, whereas hybrid approaches are more cost effective for high-intensity hazards, especially for flood mitigation along coasts (*31*) and stormwater management in cities (*32*). In other words, a combination of approaches is needed to address the drivers and impacts of climate change. Even so, the default remains engineered solutions. This is a missed opportunity to build resilience, especially in lower-income nations where dependency on natural ecosystems is high and finance for technology is limited.

Varying effectiveness and misuse of nature-based solutions

The effectiveness of nature-based solutions varies depending on the type and condition of the ecosystems, the interventions involved, how they are implemented, and the intended beneficiaries (*2*, *3*). Globally, the most cost-effective land-based mitigation potentials involve forests (especially in the tropics), followed by inland wetlands (peatland soils), coastal wetlands, and grasslands (*33*). According to the "natural climate solutions (NCS) hierarchy," protecting intact ecosystems has the highest potential for mitigation, followed by management of working lands, and lastly restoration (*34*). Preventing the loss of stored carbon by protecting ecosystems is estimated to be about twice as effective globally as restoring lost or damaged ecosystems [~4 Gt of CO_2 per year for protection compared with 2 Gt of CO_2 per year for restoration (*13*)]. Moreover, the gradual accumulation of carbon from restoration cannot compensate, in a timely way, for the rapid emission of large amounts of carbon from deforestation and degradation (*34*). Protecting intact ecosystems, especially old-growth forests, also brings the greatest benefits for biodiversity and local communities (including support for climate change adaptation) and has the lowest risk of impermanence (*35*).

The potential benefits of nature-based solutions are ultimately limited by the availability of suitable land and sea areas, and so it is the larger countries with the highest forest cover that could benefit most from their implementation (*36*). Ecosystem protection and restoration face intense competition with other uses such as coastal and urban development, production of food and fiber, and carbon-reduction technologies (e.g., wind and solar farms). However, most (85%) of the land needed for cost-effective (<US$100 per tonne) mitigation of 10 Gt of CO_2 per year comes from improving the management of existing working lands—i.e., ~2.5 billion ha of agricultural and grazing lands and production forest (*13*). Rather than displacing food and fiber production, nature-based solutions on these lands (e.g., agroforestry implementation) could enhance yields (*24*, *25*).

Even where suitable space for nature-based solutions is available, major anthropogenic stressors threaten the health and resilience of ecosystems and their ability to provide benefits to society (Fig. 1). Ecosystems are usually well adapted to natural disturbances, and some require a degree of disturbance to thrive. However, under anthropogenic climate change, the frequency and intensity of fires, floods, droughts, and invasions by pests and pathogens may be too high to allow for recovery and/or adaptation (*7*). Such climate-driven risks are already causing forest dieback across the world (*11*) and are predicted to reduce carbon sequestration by tropical forests (*37*), and many ecosystems have transitioned or are in the process of transitioning to alternative states under climate change that either cannot support human adaptation or are net sources of GHGs (*3*, *37*, *38*). Thus, unless GHG emissions are curbed, the impacts of climate change will cause the carbon stored in ecosystems to be released back into the atmosphere, accelerating warming and compromising human adaptation (Fig. 1) (*2*, *11*). Such climate-driven threats to ecosystem integrity are compounded by pollution, logging, and fragmentation (*38*), which reduce resilience via loss of biodiversity and curtail species' abilities to track shifting climate niches (*39*). Conversely, careful adaptive management of nature-based solutions over time can reduce these threats and increase resilience (*19*).

In addition, a wide range of institutional, socioeconomic, and cultural factors can hinder implementation and maintenance of nature-based solutions (Fig. 1), especially in the areas of governance and finance (*8*, *40*, *41*). Implementing nature-based solutions requires overcoming

siloed governance, as it is often a multisectoral transjurisdictional policy issue that requires cooperation within and between governments and among stakeholders with differing or conflicting priorities. Nature-based solutions are rarely implemented unless they are mainstreamed or integrated with existing planning tools (*41*). Successful implementation also requires a secure and sustainable flow of finance to the communities and projects that need it most, in the form most suitable for the local sociocultural and political context. Land ownership and access are also key considerations; there is evidence that nature-based solutions projects in Europe are more successful when implemented in designated conservation areas where land rights and access are already established (*41*). Such constraints on the feasibility of implementing nature-based solutions can reduce their potential (*36*) [e.g., the mitigation potential of restoring of degraded lands in Southeast Asia was reduced by 80% when implementation barriers were considered (*42*)].

Despite growing evidence for the multiple benefits of well-designed nature-based solutions, this term has become controversial for three broad reasons. First, nature-based solutions are being used in "greenwashing." Specifically, many of the biggest emitters (such as fossil fuel companies and the wealthy nations that subsidize them) are investing in activities labeled as nature-based solutions without also investing in robust actions to rapidly decarbonize their operations (*43*). As a result, some have dismissed nature-based solutions as a "dangerous distraction" from the urgent need to keep fossil fuels in the ground that has been co-opted by those wanting to perpetuate the "unsustainable, unjust status quo" of an energy-intensive global economy (*6*). Including nature-based solutions as carbon "offsets" while continuing business as usual not only allows for continued emissions, dismissing the need for systemic change, but also distracts attention from harmful, and often catastrophic, local impacts of fossil fuel extraction on people and the environment.

A second major concern is that actions labeled as nature-based solutions are sometimes implemented through top-down governance structures that do not respect local rights; fail to account for local voices, values, and knowledge in decision-making; and perpetuate power asymmetries (*6*). For example, there is evidence that poor-quality carbon offsets, such as plantations of single, non-native tree species, are leading to land grabs. Sometimes plantations or protected areas are established, in the name of climate change mitigation, within Indigenous peoples' territories without their knowledge or consent, ignoring their rights and cultural links with ecosystems (Fig. 3, E to G). Similar problems can arise in externally imposed adaptation strategies, which can cause maladaptation or benefit only some people at the expense of others (Fig. 3H) (*2*, *19*). Local social outcomes of nature-based solutions, such as empowerment and community cohesion, are key to ensuring that nature-based solutions can be maintained over time (*2*, *19*). Not only are human rights dependent on a healthy and safe environment, but the "recognition of human rights can itself be essential for restoring and protecting healthy ecosystems" (*44*). Indeed, Indigenous peoples and local communities often have considerable knowledge on how to form reciprocal relationships with nature and adapt to change and are playing a key role in tackling the biodiversity and climate crises (*2*, *45*). Projects that ignore their knowledge and rights are therefore both unethical and unlikely to be maintained over the long term (*2*, *45*). A related concern is that nature-based solutions are instrumentalizing cultural practices and commodifying nature, thus violating many people's values and worldviews; this issue underpins pushback against the term "nature-based solutions" by some Indigenous peoples and local community groups (*6*).

Finally, there is evidence that the misuse of nature-based solutions can harm biodiversity, which secures the flow of ecosystem functions and services and is the cornerstone of resilient, adaptive ecosystems (*2*, *46*). A key concern is that current policy and funding emphasize afforestation rather than ecosystem protection and restoration (*47*) and that investments incorrectly characterized as nature-based solutions can involve low-diversity commercial plantations of non-native species (Fig. 3, E and F). This drives biodiversity loss when plantations replace native forests, grasslands, moorlands, and peatlands (*2*, *48*). Fast-growing plantations can also compromise human adaptation by reducing water supplies (*20*). Recent work emphasizes the high risks of carbon loss from tree plantations, either because the wood is harvested for short-lived products or because the plantations are more susceptible to the impacts of fires, pests, and diseases (*2*, *47*, *48*). A paired global analysis found that plantations produce more commercially usable biomass than comparable native forests but store less carbon, have lower water availability, prevent erosion less effectively, and harbor fewer species (*49*). Plantations can also cause net harm when they distract from the imperative of effectively protecting the remaining intact ecosystems. Some governments have committed to investing in large-scale tree-planting schemes while also opening up previously protected forests for logging [e.g., (*50*)].

Avoiding the bad, harnessing the good

To address these issues, a growing number of organizations are developing evidence-based guidance on what constitutes successful, sustainable nature-based solutions (*2*). Some are framed as policy guardrails for the uptake of the concept in international climate policy (e.g., www.nbsguidelines.info); others present evaluative standards for the investment, planning, and practice of nature-based solutions [e.g., the global standard for nature-based solutions from the International Union for Conservation of Nature (IUCN) (*51*)]. All guidance converges on a set of key recommendations. In particular, there is strong agreement that nature-based solutions must be designed, implemented, and monitored by, or in close partnership with, Indigenous peoples and local communities,

Box 1. Critical research questions for evaluating nature-based solutions for climate change mitigation and adaptation.

1. How does the social and ecological effectiveness of different nature-based solutions versus alternative hybrid and technological approaches vary in space and time, as well as with the magnitude of threats from climate change?
2. How do interactions between ecosystems influence the social and ecological effectiveness and permanence of different types of nature-based solutions?
3. How and why do the different dimensions of effectiveness trade off with one another—e.g., mitigation with adaptation outcomes, social with ecological impacts, and short-term with long-term benefits?
4. How can ecosystems be adaptively and equitably governed and managed to reduce the impacts of anthropogenic stressors on the flow of multiple benefits from nature-based solutions, including biodiversity?
5. How can the multiple benefits and costs of nature-based solutions be equitably distributed within and among local stakeholders and rightsholders?
6. How do social structures and processes such as cohesion, learning, and empowerment influence the social and ecological effectiveness of nature-based solutions?
7. What forms and interplay of public and private finance can incentivize the implementation of nature-based solutions, and what are their economic and governance implications, particularly in relation to equity, permanence, and impacts on biodiversity?
8. How can the social and ecological outcomes of nature-based solutions be monitored over the long term, informed by Indigenous and local knowledge as well as mainstream science?

Mitigation Adaptation Biodiversity Livelihoods/rights

Good practice, nature-based solutions

Bolivia,
lowland tropical forest

A Indigenous land titling and forest management

(+) (+) (+) (+)

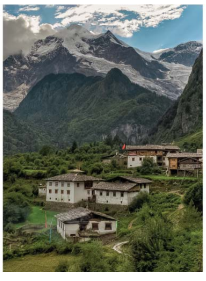
Indigenous Council of Tacana People manage 390k ha of ancestral lands
Projected 13 Mt avoided CO_2 emissions
Reported reduced slope erosion
Habitat for > 50 endangered species
Sustainable livelihoods (e.g. agroforestry)

Tanzania,
dry grasslands

B Farmer-led pasturelands management with a silvo-pastoral system

(+) (+) (+) (+)

Traditional silvo-pastoral system (Ngitili)
Protects patches of native vegetation from livestock, opening them up in dry season when feedstocks are low
Enhanced soil carbon and biodiversity likely, but not quantified
Reports of reduced soil erosion, enhanced soil fertility, and year-round food security



China,
montane forest and wetlands

C Participatory integrated watershed management

(+) (+) (+) (+)

Indigenous Naxi people self-organized into watershed management teams, empowering vulnerable groups, including women.
Protected/restored forest and wetlands, with agroforestry and sustainable fishing
Increased carbon storage likely, but not quantified
Decreased risk of erosion
Increased local incomes, food and water security
Native wildlife returned

Grenada,
coastal ecosystem

D Community-based hybrid approach to build coastal resilience

(+) (+) (+) (?)

Local residents artifically enhanced reefs and restored mangroves
Enhanced reefs reported to break wave energy and reduce erosion and flooding
Restored mangroves store carbon and protect coastline
Coral recruitment observed within 12 months
Reports of fish, octopus and lobsters on new reef structures

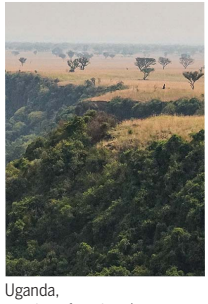
Poor social and ecological outcomes

Cambodia,
lowland tropical forest

E Exclusion of Indigenous people and afforestation with non-native monoculture

(−) (−) (−) (−)

Prey Lang forest declared a wildlife sanctuary
Increase in illegal logging as Indigenous Kuy people were denied access and could no longer patrol forest
Think Biotech granted rights to 'reforest' 34,007 ha for climate mitigation, replacing old growth forest with *Acacia* monoculture
Negative outcomes for biodiversity and climate mitigation
Further displacement and marginalization of local people

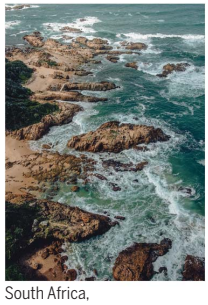


Uganda,
montane forest and savanna

F Controversial reforestation for carbon offsetting

(?) (−) (?) (−)

Mount Elgon Uganda Wildlife Authority – Forests Absorbing Carbon Emissions project marketed as a 'triple-win' by a company selling carbon credits to offset aviation emissions
Aimed to sequester 3.7 Mt CO_2e in 1994-2034 by 'reforesting' 25,000 Ha
Relied on uncompensated dispossession and violent eviction of local people
Local conflict and negative international publicity
Project abandoned after 10 years

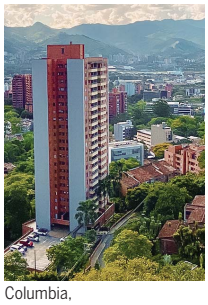


South Africa,
marine ecosystems

G Top-down implementation of marine protected areas

(?) (−) (+) (−)

Dwesa-Cwebe Marine Protected Area forced racially-based relocations
Criminalization of customs, violent suppression of illegal harvesting
Traditional ecological knowledge ignored, cultural heritage eroded
Loss of tenure rights led to loss of food security and livelihoods



Columbia,
urban and montane ecosystems

H Peri-urban land grab for greenbelt development

(+) (?) (+) (−)

Greenbelt project established around the city of Medellin
Protected natural habitats
Enhanced carbon storage and biodiversity likely, but not quantified
Decreased risk of landslides
Displaced vulnerable communities, but not high-income ones
Social tensions from lack of local consultation and participation, poor planning, and green gentrification

PHOTOS: (A) AMY ROLLO/UNSPLASH, (B) MARIOLA GROBELSKA/UNSPLASH, (C) JAMES WHEELER/UNSPLASH, (D) BISHNU SARANGI/PIXABAY, (E) YOUNG SHIH/UNSPLASH, (F) JULIE RICARD/UNSPLASH, (G) THOMAS PEHAM/UNSPLASH, (H) MIKE SWIGUNSKI/UNSPLASH

Fig. 3. Examples of projects that meet guidelines for successful, sustainable nature-based solutions and those that do not. To be categorized as a nature-based solution, an intervention must address at least one societal challenge while also bringing local benefits and supporting biodiversity. Therefore, although bioenergy with carbon capture and storage (commonly referred to as BECCS) and afforestation (i.e., planting trees where they do not naturally occur) are biological approaches to carbon sequestration, they are not nature-based solutions, nor are protected areas established without regard for the rights of local communities (*2*). Shown here are good-practice examples (**A** to **D**) in a range of ecosystems involving a variety of nature-based solutions that, as far as can be determined from available information, meet guidelines for successful, sustainable nature-based solutions (www.nbsguidelines.info). By contrast, examples of projects established for mitigating climate change (**E** and **F**), protecting biodiversity (**G**), or reducing flood and landslide risk (**H**) have poor social and/or ecological outcomes and thus do not qualify as nature-based solutions. Symbols denote mitigation, adaptation, biodiversity, and livelihoods or human rights outcomes reported as positive (+) or negative (−) (or likely to be positive or negative on the basis of current evidence). Question marks denote outcomes that are unknown or difficult to predict. Note that a question mark for adaptation is assigned to the scenario detailed in (H) because only one dimension of adaptation was addressed [reduction of exposure to the immediate climate change impact (in this case, landslides)], whereas other dimensions were compromised. For more information about (A) to (D) and other examples of good-practice nature-based solutions, see https://casestudies.naturebasedsolutionsinitiative.org/. Information about projects (E) to (H) can be found in (*56*–*59*).

reinforcing local rights and delivering benefits for local people. However, more clarity is needed on how the concept of nature-based solutions can align with Indigenous and local values and worldviews and, in particular, to explain that such solutions should not represent a commodification of nature (*6*). Nature-based solutions need to be understood as ways of working with, and as part of, nature and framed to ensure that multiple values of nature are respected. For this, the nature-based solutions community could learn from the Intergovernmental Platform on Biodiversity and Ecosystem Services in its efforts to develop an inclusive framework for understanding how contributions from nature affect people both positively and negatively (*52*).

There is also consensus on the critical importance of ensuring that nature-based solutions support biodiversity by protecting, restoring, and connecting a wide range of native habitats across landscapes and seascapes and by monitoring ecological outcomes (which are rarely quantified at present). Guidelines could more effectively promote biodiversity by recommending the use of the natural climate solutions hierarchy that prioritizes protection of intact ecosystems (*34*), as well as approaches that allow ecosystems to reach their full potential with minimal intervention [i.e., through "proforestation" (*53*)]. Practitioners would also benefit from understanding impermanence risks and adopting adaptive approaches to restoration. Given that local climate and disturbance regimes are changing, in some locations it may be necessary to restore and manage ecosystems by using different species that are able to thrive under unfamiliar conditions, including through assisted migration (*54*).

Guidelines are also clear that nature-based solutions are not an alternative to keeping fossil fuels in the ground. The challenge is how to direct rapidly growing public- and private-sector finance (especially via booming voluntary carbon markets) toward high-integrity nature-based solutions projects that are well planned and do not delay decarbonization. Part of the solution is regulation that restricts investment in nature-based solutions–related offsets to those organizations with ambitious yet robust and verifiable action plans to rapidly phase out the use of fossil fuels. This includes meeting stringent criteria for companies to reduce emissions throughout their operations and supply chains to be in line with the Paris Agreement's 1.5°C increase limit before or in addition to investing in robust nature-based solutions—for example, by adhering to the plan of the Science Based Targets initiative (https://sciencebasedtargets.org/). To ensure compliance, nature-based solutions should be rigorously assessed and validated, by methods such as long-term monitoring of social and ecological impacts. Government policy that supports robust accountability and regulatory frameworks for nature-based solutions would be transformative (*2*).

Ensuring the long-term social and ecological integrity of nature-based solutions requires an improved evidence base, informed by scientific, practitioner, and local and Indigenous knowledge. There is an urgent need for enhanced understanding of where, when, how, and for whom nature-based solutions can support both mitigation and adaptation, especially in marine and nonforest ecosystems, low-income nations in general, and their cities in particular (*20*, *21*). This research needs to take a holistic approach, for example, by considering how nature-based solutions influence the multiple dimensions of adaptation, not just exposure to immediate climate change impacts, and by comparing the benefits and costs of nature-based solutions with hybrid and technological alternatives (Box 1).

More regional and national scenarios and sectoral models of nature-based solutions climate change mitigation potentials are urgently required. These must be firmly grounded in the local policy and socioeconomic and cultural context, with robust consideration of permanence and leakage risks, interactions between ecosystems, feasibility, and outcomes for biodiversity, local people, and the economy (*36*). Such models must also distinguish afforestation from reforestation (*47*–*49*). Integrating landscape-scale experiments with recent advances in multispectral remote sensing, large-scale ecological observation networks, disturbance ecology, and mechanistic vegetation modeling will improve quantification of risks to the stability of ecosystems (*11*). Such research would also support the development of high-quality metrics that capture the multidimensional nature of biodiversity as well as the social outcomes of interventions. This information will allow baselines to be established and effects of nature-based solutions to be mapped and monitored over time, which is critical to improving adaptive management. Meanwhile, new research indicates that participation in nature-based solutions can lead to more-sustainable lifestyle choices (*55*); the role of nature-based solutions in enabling system change is a rich area for future study.

Addressing these knowledge gaps involves genuine collaboration between social and natural scientists, as well as between the scientific community and those Indigenous peoples and local communities who have been working with nature, as part of nature, for millennia. Only through such inter- and transdisciplinary efforts will we be able to implement nature-based solutions in a way that accounts for their multiple values. Such efforts are urgently needed to channel growing private and public climate finance to the projects and people that need it most—and in a form that will build rather than compromise the health and resilience of the social-ecological systems involved.

Conclusion

Nature-based solutions can make an important contribution to reaching net-zero carbon emissions this century, but only if combined with other climate solutions, including substantial cuts in GHG emissions across all sectors of the economy. This statement is not an argument against scaling up nature-based solutions. Instead, it underscores the need to consider the many other well-evidenced benefits of nature-based solutions, especially their critical role supporting social and ecological adaptation to climate change. Achieving net-zero carbon emissions and transitioning to a nature-positive economy will also require systemic change in the way we behave as societies, shifting to a dominant worldview that is based on valuing quality of life and human well-being rather than material wealth—and connection with nature rather than its conquest. Signals such as the rise of climate and nature grassroots activism indicate that this shift is taking place. If carefully implemented to ensure that multiple values of the natural world are respected, nature-based solutions offer an opportunity to accelerate this transition while also slowing warming, building resilience, and protecting biodiversity.

REFERENCES AND NOTES

1. E. Cohen-Shacham *et al*., *Environ. Sci. Policy* **98**, 20–29 (2019).
2. N. Seddon *et al*., *Glob. Change Biol.* **27**, 1518–1546 (2021).
3. N. Seddon *et al*., *Philos. Trans. R. Soc. B* **375**, 20190120 (2020).
4. S. Díaz *et al*., *Science* **359**, 270–272 (2018).
5. United Nations Framework Convention on Climate Change (UNFCCC), Glasgow Climate Pact. Decision text from CoP26 (2021); https://unfccc.int/sites/default/files/resource/cp2021_12_add1E.pdf.
6. M. S. Melanidis, S. Hagerman, *Environ. Sci. Policy* **132**, 273–281 (2022).
7. Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, P. R. Shukla *et al*., Eds. (Cambridge Univ. Press, 2022).
8. C. J. Nolan, C. B. Field, K. J. Mach, *Nat. Rev. Earth Environ.* **2**, 436–446 (2021).
9. J. Howard *et al*., *Front. Ecol. Environ.* **15**, 42–50 (2017).
10. H. D. Matthews *et al*., *Commun. Earth Environ.* **3**, 65 (2022).
11. W. R. L. Anderegg *et al*., *Science* **368**, eaaz7005 (2020).
12. A. C. Soterroni *et al*., *Sci. Adv.* **5**, eaav7336 (2019).
13. C. A. J. Girardin *et al*., *Nature* **593**, 191–194 (2021).
14. C. M. Anderson *et al*., *Science* **363**, 933–934 (2019).
15. UNFCCC, Paris Agreement (United Nations, 2015); https://unfccc.int/sites/default/files/english_paris_agreement.pdf.
16. L. Berrang-Ford *et al*., *Nat. Clim. Change* **11**, 989–1000 (2021).
17. United Nations Environment Programme, "Adaptation Gap Report 2021: The gathering storm – Adapting to climate change in a post-pandemic world" (2021); https://www.unep.org/resources/adaptation-gap-report-2021.
18. I. Palomo *et al*., *One Earth* **4**, 730–741 (2021).
19. T. Devisscher, J. Spies, V. C. Griessa, *Land Use Policy* **103**, 105317 (2021).
20. A. Chausson *et al*., *Glob. Change Biol.* **26**, 6134–6155 (2020).
21. K. Sudmeier-Rieux *et al*., *Nat. Sustain.* **4**, 803–810 (2021).
22. L. Ruangpan *et al*., *Nat. Hazards Earth Syst. Sci.* **20**, 243–270 (2020).

23. P. Herath, M. Thatcher, H. Jin, X. Bai, *Sustain Cities Soc.* **72**, 103072 (2021).
24. S. Kuyah *et al*., *Agron. Sustain. Dev.* **39**, 47 (2019).
25. S. Kay *et al*., *Land Use Policy* **83**, 581–593 (2019).
26. The Royal Society, Royal Academy of Engineering, "Greenhouse gas removal" (The Royal Society, 2018); https://royalsociety.org/greenhouse-gas-removal.
27. J. Hinkel *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3292–3297 (2014).
28. A. B. Rodriguez *et al*., *Nat. Clim. Chang.* **4**, 493–497 (2014).
29. B. P. Harman, S. Heyenga, B. M. Taylor, C. S. Fletcher, *J. Coast. Res.* **31**, 790–801 (2015).
30. S. Hynes, R. Burger, J. Tudella, D. Norton, W. Chen, *Ecol. Econ.* **194**, 107349 (2022).
31. R. L. Morris, A. Boxshall, S. E. Swearer, *Nat. Clim. Chang.* **10**, 485–487 (2020).
32. W. Chen *et al*., *Int. J. Disaster Risk Reduct.* **54**, 102045 (2021).
33. J. Reise *et al*., "Nature-based solutions and global climate protection. Assessment of their global mitigation potential and recommendations for international climate policy" (German Environment Agency, 2022).
34. S. C. Cook-Patton *et al*., *Nat. Clim. Chang.* **11**, 1027–1034 (2021).
35. J. E. M. Watson *et al*., *Nat. Ecol. Evol.* **2**, 599–610 (2018).
36. S. Roe *et al*., *Glob. Change Biol.* **27**, 6025–6058 (2021).
37. L. V. Gatti *et al*., *Nature* **595**, 388–393 (2021).
38. C. A. Boulton, T. M. Lenton, N. Boers, *Nat. Clim. Chang.* **12**, 271–278 (2022).
39. C. Moritz, R. Agudo, *Science* **341**, 504–508 (2013).
40. I. Schulte, J. Eggers, J. Ø. Nielsen, S. Fuss, *Environ. Res. Lett.* **17**, 013002 (2021).
41. J. Nalau, S. Becken, B. Mackey, *Environ. Sci. Policy* **89**, 357–364 (2018).
42. Y. Zeng *et al*., *Nat. Clim. Chang.* **10**, 842–844 (2020).
43. M. Li, G. Trencher, J. Asuka, *PLOS ONE* **17**, e0263596 (2022).
44. H. Tugendhat, "Re-thinking nature-based solutions: seeking transformative change through culture and rights" (Forest Peoples Programme, 2020).
45. J. D. Ford *et al*., *One Earth* **2**, 532–543 (2020).
46. T. H. Oliver *et al*., *Trends Ecol. Evol.* **30**, 673–684 (2015).
47. S. L. Lewis, C. E. Wheeler, E. T. A. Mitchard, A. Koch, *Nature* **568**, 25–28 (2019).
48. F. Fleischman *et al*., *Bioscience* **70**, 947–950 (2020).
49. F. Hua *et al*., *Science* **376**, 839–844 (2022).
50. R. Frazin, "Trump creates federal government council on global tree planting initiative," *The Hill* (2020); https://thehill.com/policy/energy-environment/520852-trump-creates-federal-government-council-on-global-tree-planting.
51. International Union for Conservation of Nature (IUCN), *Guidance for using the IUCN Global Standard for nature-based solutions. A user-friendly framework for the verification, design and scaling up or nature-based solutions* (IUCN, ed. 1, 2020); https://portals.iucn.org/library/sites/library/files/documents/2020-021-En.pdf.
52. U. Pascual *et al*., *Curr. Opin. Environ. Sustain.* **26–27**, 7–16 (2017).
53. W. R. Moomaw, S. A. Masino, E. K. Faison, *Front. For. Glob. Change* **2**, 27 (2019).
54. G. Peterson St-Laurent, L. E. Oakes, M. Cross, S. Hagerman, *Commun. Biol.* **4**, 39 (2021).
55. M. L. Cárdenas *et al*., *Sustainability* **13**, 4344 (2021).
56. A. Scheidel, C. Work, *Land Use Policy* **77**, 9–18 (2018).
57. C. Cavanagh, T. A. Benjaminsen, *Geoforum* **56**, 55–65 (2014).
58. M. Sowman, J. Sunde, *Ocean Coast. Manage.* **157**, 168–179 (2018).
59. I. Anguelovski, C. Irazábal-Zurita, J. J. Connolly, *Int. J. Urban Reg. Res.* **43**, 133–156 (2019).

ACKNOWLEDGMENTS

The author thanks A. Chausson, J. Lynch, A. Smith, A. Soterroni, B. Turner, and two anonymous reviewers for valuable comments and A. Wagner and D. Seddon for help with the figures. **Funding:** This work was supported by the Waterloo Foundation and the Natural Environment Research Council. **Competing interests:** The author is on the Adaptation Committee of the UK's Climate Change Committee, advising on land, biodiversity, and agriculture.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 31 March 2022; accepted 18 May 2022
10.1126/science.abn9668

REVIEW

Climate change and the urgency to transform food systems

Monika Zurek[1]*, Aniek Hebinck[2], Odirilwe Selomane[3]

Without rapid changes to agriculture and food systems, the goals of the 2015 Paris Agreement on climate change will not be met. Food systems are one of the most important contributors to greenhouse gas (GHG) emissions, but they also need to be adapted to cope with climate change impacts. Although many options exist to reduce GHG emissions in the food system, efforts to develop implementable transformation pathways are hampered by a combination of structural challenges such as fragmented decision-making, vested interests, and power imbalances in the climate policy and food communities, all of which are compounded by a lack of joint vision. New processes and governance arrangements are urgently needed for dealing with potential trade-offs among mitigation options and their food security implications.

Climate change poses one of the greatest threats to human societies, demanding immediate and coordinated actions across all sectors (*1*). Food systems are one of the most important contributors to climate change (*2*) and could compromise efforts to achieve the 2015 Paris Agreement targets (*3*). At the same time, food systems themselves will also need to further adapt to climate change impacts. The latest Intergovernmental Panel on Climate Change report shows that climate change has already negatively affected food production across the world and contributed to malnutrition (*4*). Temperature rises beyond 1.5°C are expected to transform terrestrial land ecosystems and shift climate zones (*5*), pressuring food security and livelihoods by affecting the productivity of crops and livestock (*4*), and warming of the oceans will reduce the productivity of fisheries and aquaculture (*6*). Together with more extreme weather events and sea-level rise, this level of temperature increase will exacerbate inequities in food access as food prices increase (*7*).

Simultaneously, food systems are responsible for about one-third of global anthropogenic greenhouse gas (GHG) emissions (*2*), presenting a major challenge—but also opportunities—for climate change mitigation (*8*). There are three major pathways through which the food system contributes to GHG emissions that present entry points for transformative change. The first pathway is through crop and livestock production, including all of the activities required to ensure that raw products leave the farm. These activities generate GHG emissions mainly through the methane and nitrous oxide produced from enteric fermentation by domestic ruminants (cows, sheep, and goats) and their manure, synthetic fertilizer applications on crops, and methane production from flooded rice fields (*9*, *10*). Together, crops and livestock systems currently contribute 10 to 14% of total GHG emissions, which could increase to 40% by 2050 under some scenarios (*7*, *9*). The second major pathway is land-use change, which contributes to GHG emissions mainly through deforestation and destruction of peatlands for agricultural purposes. Agriculture-related land-use emissions are estimated to be between 5 and 14% of total emissions (*7*, *9*). The third pathway is through food-related activities beyond the farm gate, ranging from food processing and transport to food consumption. Food system–related emissions beyond the farmgate are estimated to be between 5 and 10% of total emissions (*9*).

"Transformation of food systems around the globe is urgent..."

Coordinated and successful implementation of a "menu" of mitigation and adaptation options for agriculture and food systems on a global scale could reduce GHG emissions to a safe level and support transformation to sustainable food systems (*10*). Mitigation options in food systems are generally organized around four key areas: improvements in the management of crops and livestock, land-use change, and food value chains, as well as altering food consumption patterns and reducing food waste. Although agricultural activities and land-use change are leading to a higher proportion of food system emissions than post-farm-gate activities, consumer dietary choices are a substantial factor driving decisions made on the farm. That said, post-farm emissions, including those from the energy sources used in food processing, food transport, food storage, and cooking, have been rising substantially in recent years, requiring a rethink of mitigation strategies for the food sector (*11*). A look across the whole food system therefore becomes important for finding the biggest levers of change.

[1]Food Systems Transformation Group, Environmental Change Institute, University of Oxford, Oxford, UK. [2]Dutch Research Institute for Transitions (DRIFT), Erasmus University Rotterdam, Rotterdam, Netherlands. [3]Centre for Sustainability Transitions, Stellenbosch University, Stellenbosch, South Africa.
*Corresponding author. Email: monika.zurek@eci.ox.ac.uk

Designing a menu of mitigation (and adaptation) measures

Although mitigation and adaptation options are plentiful in the food system (*12*), their implementation remains fragmented and uncoordinated, risking trade-offs with other food system outcomes such as food security or livelihoods (*13*, *14*). Harnessing the climate change mitigation and adaptation potential in the food system will require a critical systems perspective (*15*) to understand the pros and cons of these options, as well as relations among the different actors that might affect the implementation of an intervention. This approach could also help us to see beyond mitigation and adaptation options that target agriculture, which until now has been a dominant focus in literature and practice (*8*), to better include other activities down the food chain. For this, the diverse and interrelated activities that make up the food chain, as well as the wider social-ecological context and driving forces within it, need to be considered (*15*). A food system approach (Fig. 1) can illuminate several intervention points along the chain by modifying the drivers of food system activities (*16*). For example, farm activities can either contribute to GHG emissions or sequester carbon, depending on production practices (*17*). Similarly, dietary choices at the household level could significantly influence meat-production-related emissions (*18*).

Transformation of food systems around the globe is urgent, not only because of their GHG emissions but also because they fall completely short in equitably distributing food and providing food and nutrition security (*19*), resulting in hunger, malnutrition, and overconsumption (*20*). In addition, their wider environmental footprint related to biodiversity loss, deforestation, soil degradation, and water pollution is a key driver of environmental degradation (*21*). As currently organized, the food system also falls short on providing equal economic opportunities to food system actors or social equity at large (*15*, *22*) (Fig. 1).

Here, we review key food system climate change mitigation options and take a systems perspective to explore interactions with the main food system outcomes. We then examine some of the key stumbling blocks to achieving the necessary mitigation efforts in food systems and point to new ideas for overcoming these to bring about tangible food system change with mitigation benefits.

Climate change mitigation options in food systems

Climate change mitigation strategies across the entire food system fall into four main categories. We review some of the most important options, which range from improvements to cropping

systems, livestock production, and supply chain activities to changing demand for products high in GHG emissions (Table 1). Although all of these options have GHG emission reduction benefits, they also have implications for other food system outcomes by creating potential synergies and trade-offs. Table 1 presents a few examples of these interactions, which need to be analyzed in their specific contexts to assess their true benefits.

Fostering synergies can offer a multitude of co-benefits. For example, agroforestry practices have major benefits, not only for the environment by fixing nitrogen and enhancing soil carbon sequestration, but also for society by increasing crop productivity by enhancing soil nutrients and organic matter. However, the adoption of agroforestry practices is often complicated by obstacles specific to the farming system context. Managing the risk of an unsuccessful transition can be challenging when no economic fallback mechanisms are available or when different inputs or new knowledge to change practices is required.

Adoption of a mitigation measure can also be complicated by associated trade-offs with food system outcomes. For example, closing yield gaps (i.e., the difference between the potential yield for a particular crop under optimal conditions and average farmers' yields in a particular location) is important for climate mitigation through reduction of demand for new land. However, closing the yield gap requires resource inputs, such as water and fertilizer (*23*), which might not be locally available, and their use could have other, possibly negative environmental impacts (*24*). Similarly, strategies for reducing GHG emissions from enteric fermentation in ruminants, which produces methane and is the biggest contributor to food-related GHG emissions, for example by incorporating marine algae into cattle feed, has a large mitigation potential, but large-scale seaweed harvesting will likely result in negative effects on marine ecosystems and also reduce their carbon sequestration potential (*30*).

Although options for mitigating climate change in agriculture and the wider food system exist, it is still not clear how these can be combined to reduce GHG emissions sufficiently to meet the 2015 Paris Agreement. Wollenberg *et al.* (*41*) were the first to calculate a potential target for agricultural emissions reductions by 2030 so that the sector could feasibly aim for the 2°C goal of the Agreement. The authors then investigated whether this target could be reached using various currently available agricultural mitigation technologies (for examples, see Table 1) and concluded that they would only deliver 21 to 40% of the required reductions in GHG emissions. Far-reaching efforts to develop further transformational technological and policy options are essential if agriculture and food systems are to play a significant role in achieving the Paris Agreement goals.

Systemic challenges to reducing food-related emissions

Despite scientific validation of possible mitigation options and strategies for addressing climate change, society has collectively failed to implement options or to "bend the curve" toward lower emissions. Several systemic and power-related issues, such as lack of coordinated climate action and vested interests in fossil fuel industries, are partially to blame, as are unbalanced power relations within food systems, which hinder progress in the adoption of mitigation strategies (*42*, *43*). In addition to obstructing reduction of emissions, these issues have also exacerbated inequities between high- and middle- and low-income countries, because climate impacts have thus far mostly affected the latter (*20*). Stoddard *et al.* (*42*) set out the structural issues that have prevented any bending of the curve, and in this section we will also explore how these structural issues result in trade-offs between equity and climate change mitigation in food systems.

Addressing climate change is to a large extent a political matter that involves negotiations between governments and other stakeholders about how to coordinate action on a global scale (*44*). By coordinating and giving shape to climate action, the so-called "climate change regime" (*42*, *45*, *46*) has provided the dominant structure for such negotiations. It is this dominant governance arrangement that also decides about the allocation of responsibility for mitigation of climate change between states amidst north-south geopolitics (*44*). However, this is not an equal playing field, and concerns have been raised about how powerful countries have undue influence on the process. Furthermore, hindering coordinated action in this governance arrangement are the lack of binding targets for emissions, requirements for technology transfer, and funding for low-income countries (*42*). These issues directly influence food-related mitigation strategies in low- and middle-income countries that are the most vulnerable to climate impacts (*13*). In addition, the consensus-based decision-making approach might block the approval of mitigation options that benefit low- to middle-income countries but disadvantage high-income countries (*42*).

Although global climate governance has developed targets for the reduction of fossil fuel use and GHG emissions at a national level, wealthy countries nevertheless often outsource their environmental footprint and GHG emissions to other, poorer countries while also using land elsewhere for carbon offsetting at the expense of the people who live there (*47*, *48*). In addition,

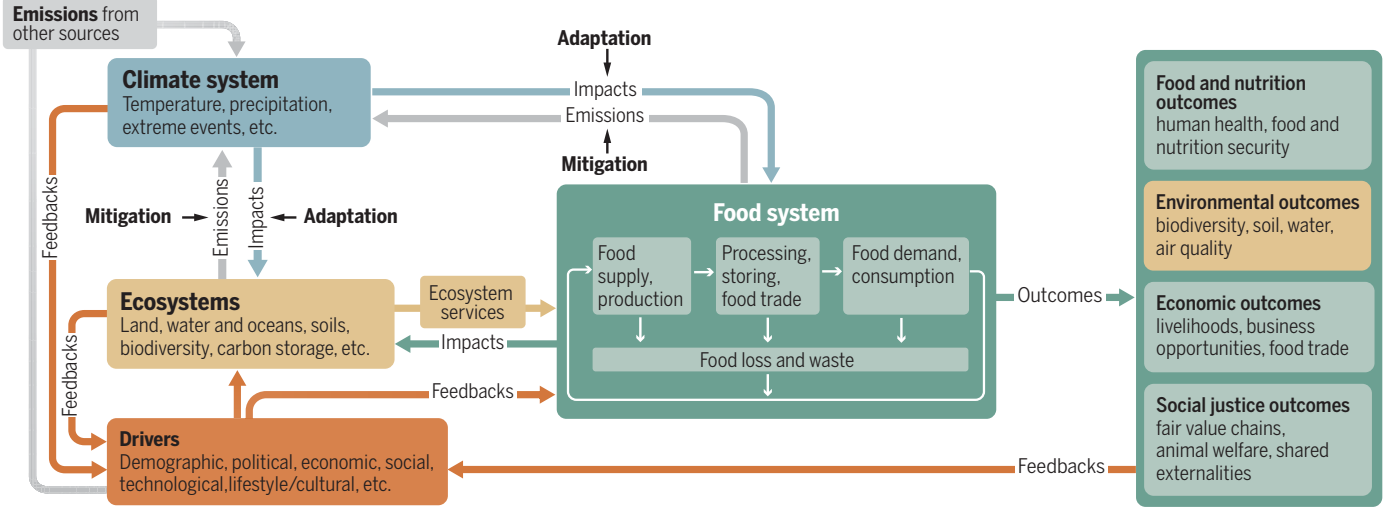


Fig. 1. How climate change, ecosystems, and food system drivers interact to affect food security [adapted from (*9*)].

Table 1. Climate change mitigation strategies across the food system.

MITIGATION AREAS	FOOD SYSTEM RESPONSES	EXAMPLES OF POTENTIAL INTERACTIONS WITH OTHER FOOD SYSTEM OUTCOMES (FOOD AND NUTRITION SECURITY, ECONOMIC, ENVIRONMENTAL, AND SOCIAL)
Improved crop management	Reducing nitrous oxide emissions from synthetic fertilizer applications	Synthetic fertilizer applications are important to current food systems, especially because manure and legumes can only provide a portion of total nitrogen demands of crop production. They have contributed to substantial gains in productivity of food crops and will continue to be important going forward, because demand for food is expected to increase. However, overapplication of fertilizer has led to major environmental impacts (*21*).
	Reducing methane emissions from paddy rice	Implementation of new agricultural management practices (e.g., alternative wetting and drying) by the many (small holder) farmers globally requires massive input from extension services (*25*), which brings uncertainties about the effectiveness of implementation.
	Improving land-use management for carbon sequestration (and reducing its losses)	The potential of carbon sequestration in agricultural lands is debated (e.g. issues with permanence), although it could be (with regional and local variations) considered a co-benefit of improving cropland and grazing land management (*26*). Restoration of peatlands and the reforestation of marginal and unimprovable agricultural lands should be a priority but conflicts with the increased demand for food (*27*).
	Closing yield gaps (differences between yields under optimal conditions and those attainable in farmers' fields)	Yield gap reduction has a substantial role to play in reducing the land needed for food. Improving yield gaps relies primarily on nutrients (fertilizer) and water management (*23*). In some areas, water required to close the yield gap might not be locally available (*24*). In terms of nutrients, some areas and regions of the world such as sub-Saharan Africa will need to increase their fertilizer use (*28*), and this will further increase GHG emissions. By contrast, many other parts of the world need to reduce fertilizer overapplication.
	Using agroforestry	Agroforestry is a land-sharing strategy that accommodates both agricultural production and biodiversity protection, resulting in improved nitrogen fixation, land and ecosystem health, and soil carbon sequestration, among other benefits. However, the implementation of this strategy depends on landowners and managers accepting and adopting these practices as well as various other socioeconomic barriers. Agroforestry suffers from similar challenges as conservation agriculture (*29*). To be successful, it will require investment to facilitate uptake in a way that is beneficial to landowners and managers.
Improved livestock management	Using better grazing land management	
	Improving manure management	
	Using higher-quality feed	There are various new options for reducing methane emissions by changing feeding practices for ruminants. For example, Roque *et al*. (*30*) suggested that introducing seaweed into the diets of cattle can reduce their methane emissions by up to 80% by changing the bacterial community composition in their guts. However, increasing the scale of seaweed harvesting would have large implications for marine ecosystems, including their carbon sequestration potential.
	Reducing enteric fermentation	Two main strategies for reducing enteric fermentation include feed additives and improving feed digestibility. Feed additives, while reducing GHG emissions from livestock, can leave toxic residues and have independent environmental impacts (*31*). Given the increasing risks from toxic residues and antibiotic and pesticide resistance, feed additives are not a clear way forward for mitigation (*32*).
	Reducing nitrous oxide through manure management	
	Sequestering carbon in pastures	
	Implementing best animal husbandry and management practices	
	Using nonanimal protein sources	
	Using microbial protein as feedstuff	
Improved supply chain	Improving food transport and distribution	Mitigation options here take two general forms. In low- and middle-income countries, where storage and processing facilities may be lacking, mitigation is geared toward reducing food loss through innovations and technology (e.g., cool storage options (*34*). In upper middle– and high-income countries, where use of technology and infrastructure is widespread, mitigation is geared toward improving energy use efficiency and transitioning toward renewable energy sources (*8*). A potential trade-off with the food system outcomes depends on the type of renewable energy sources used, e.g., the potential impacts of biofuels on food security is well documented (*35*).
	Improving efficiency and sustainability of food processing, retail, and agrifood industries	
	Improving energy efficiencies of agriculture	

Continued on next page

MITIGATION AREAS	FOOD SYSTEM RESPONSES	EXAMPLES OF POTENTIAL INTERACTIONS WITH OTHER FOOD SYSTEM OUTCOMES (FOOD AND NUTRITION SECURITY, ECONOMIC, ENVIRONMENTAL, AND SOCIAL)
	Reducing food loss	A large share of the food produced is never consumed. Reducing food loss would allow smaller yields to meet global food demand and also reduce emissions. Mitigation measures to address food loss often come in the form of innovations to improve the efficiency of food harvesting and processing. These innovations need to be both accessible and affordable to reach middle and smallholder farmers (*10*, *34*).
Demand management	Making dietary changes toward sustainable consumption and healthy diets	Reduction in meat (especially beef and lamb) consumption is expected to have the biggest outcome for climate change and the environment (*9*, *33*), especially as food demand, and especially for meat, is projected to increase. A growing number of upper- and middle-income consumers overconsume food, contributing to food demand, GHG emissions, and food waste (*33*, *36*). Switching to healthy diets and following food guidelines has the potential to improve environmental sustainability and mitigation of climate change and also improve health outcomes (*37*, *38*). However, healthy diets may be unaffordable and/or inaccessible to most of the world's poor and marginalized (*39*).
	Reducing food waste	The Food Waste Index report estimates that nearly a billion metric tons of food was wasted in 2019. More than 60% of this waste was due to household waste, with food service and retail contributing 26 and 13%, respectively (*40*). Reducing food waste has multiple co-benefits and provides synergistic outcomes for people by improving food security and regulating prices and for the planet by reducing pressure on land, biodiversity, and climate change.

vested interests from fossil fuel and dependent industries intervene on the decision-making processes at the national level (*42*), often pushing for nontransformative solutions that include technological optimism (*49*). Influence from such vested interests tends to lead to the adoption of the least disruptive changes, often leaning on future technological breakthroughs to ultimately justify business as usual (*49*, *50*). Similarly, in the debate on the role of food systems in climate change mitigation, more narrow technological solutions have received much attention, often overlooking unintended or hidden social justice consequences (*51*, *52*), for example, when innovative solutions are inaccessible to vulnerable food system actors, thus further widening the gap between rich and poor (*13*). In addition, the focus on tech-based solutions has sidelined the debate on the role of alternative, innovative solutions such as agro-ecology that are more locally based and scale appropriate for small-scale farmers (*50*).

The financial sector has also played an important role in shaping today's food economy by funding fossil fuel–reliant industries and business practices. Large-scale industrial agricultural practices are made possible by private sector finance and investments (*53*). By investing in responses and solutions that are preferred by the financial sector, financial actors have been influential in determining how climate risks are managed by the food system, for example, by proposing various carbon-trading mechanisms (*54*). However, because of the close relation between fossil fuel dependence and economic growth, a conflict arises between the feasibility of combining the current economic growth paradigm with the successful implementation of climate change mitigation options (*55*). Calls for more transformative and just climate governance require the "polluter elite" and trans-national companies to take responsibility (*43*). In this scenario, financial capital institutions will have to develop and implement more innovative finance mechanisms to support transformative food system practices (*56*).

Although justice has been a core motivation for governing climate change at the global level, it is primarily understood in terms of equitable responsibility for or responses to climate change. This notion assumes that nation-states can protect and enforce climate justice, which overlooks multiple dimensions and narrows down action to north-south or developed-developing divides (*43*). This interpretation has obstructed the implementation of more transformative mitigation options. First, the focus on the nation-state has overlooked the role of private sector actors, such as the food industry, in climate mitigation. Whereas innovation of alternative and more sustainable practices and products is considered key for climate mitigation, decreased public spending has left their development primarily in the hands of private actors who often pursue for-profit aims rather than the common good (*13*). Second, much of the debate on climate mitigation is viewed from a global scale, disconnected from the local level and the consequences that mitigation measures might have on small communities and individuals (*57*, *58*). Not ignoring potential incompatibilities, even if a compromise could be found between globally defined mitigation options and locally oriented adaptation measures, these would be difficult to implement (*59*).

Interlocking decisions

Food systems, particularly agricultural production, are caught between the need to reduce GHG emissions and the need to adapt to new temperature regimes, precipitation patterns, and extreme future events. Policy and decision makers are faced with three interlocking decisions about mitigation of climate change in food systems: (i) available options in their specific contexts; (ii) how much GHG emission reduction can be achieved by each measure or a combination of measures; and (iii) how these options interact with food and nutrition security, the economic and social outcomes of the food system, and necessary climate adaptation measures. Table 1 shows examples of the connections between a range of mitigation options and food system outcomes. Reducing the amount of ruminant meat in diets, for example, has been discussed in many countries as the main avenue for consumers to contribute to GHG emission reductions. This can also help with the unwelcome negative health impacts of meat overconsumption but will affect livestock farmers' livelihoods directly by reducing demand and will also have implications for land-use and landscape management and, thus, biodiversity.

Navigating these questions will require evidence on the pros and cons for each mitigation option within a specific food system, and eventually decisions must be made that will very likely not please everyone. Because of the interconnected nature of food systems,

these choices will bring with them unintended and unanticipated consequences, resulting in trade-offs for some actors within the system (e.g., reducing meat consumption will change livestock producers' income) and trade-offs between food system outcomes (e.g., environmental footprint versus income). Making these trade-offs more visible is important for finding ways to address them from the start. This is where a food system perspective is essential because it connects the activities of various stakeholders to food system outcomes. To change outcomes, activities need to change, incentivized by changing food system drivers (e.g., governance, institutional structures, tax regimes, and available science and technology options).

Reforming current structures and vested interests

How decisions are made and who makes them depends very much on the specific context of a food system and how it is governed. Many governments are now including agricultural mitigation options in their Nationally Determined Contributions to the Paris Agreement (NDCs), which are then meant to translate into concrete incentives for farmers and/or food industry actors to implement mitigation actions. What input these actors have when the decisions are taken depends on the institutional mechanisms available within their respective countries. Although many technical options are becoming available for both adaptation and mitigation in different food system contexts, it has not been straightforward to translate technologies into tangible changes on the ground. Even more difficult is balancing these options with the main task of providing food and nutrition security while also providing livelihoods and economic opportunities, managing wider ecosystem outcomes, and making food systems more equitable. This is where understanding vested interests on one side and the current structures governing climate change policy making and the entire food system becomes important, especially if we want a just transition of our economy and our food systems that includes vulnerable groups.

Although food systems differ greatly across the world in their components and contexts, we will nevertheless need to develop clear transformation pathways to achieve the Paris Agreement goals. How to decide on the right mixture of mitigation and adaptation options requires positive visions specifying what food system outcomes the actors want and what trade-offs they are willing to make. This requires negotiation between actors to provide coordinated innovation pathways and trade-off management. Several steps could achieve this. First, creating a map, ideally with the relevant stakeholders, of the particular food system with its dynamics, actors, activities, and outcomes helps to create a joint understanding of the process and boundaries involved [for an example of a food system map, see (*60*)]. Ideally this should include a set of compatible, integrated food system outcome metrics (*15*). The food system map should be made at the scale at which decisions on mitigation actions are taken (e.g., at a national level for the NDCs). A second step would be using participatory foresight methods, such as scenario planning or visioning, and the food system map so that food system scenarios can be built that explore the implications and trade-offs of possible mitigation and adaptation options. Here, it is particularly important to engage multiple stakeholders across the food system and include vulnerable groups and actors who might be negatively affected by possible changes. On the basis of the scenario analysis, coordinated, systemic mitigation pathways that include various options for change can be developed for the whole system. Tailored translation of mitigation pathways should be developed into actions for different food system actors, such as producers, value chain actors, consumers, or policy makers. This step needs to reduce or deal with trade-offs/unintended consequences. Evaluation and monitoring of outcomes based on selected food system metrics are needed to determine whether and how well the implemented mitigation pathways work. Finally, actions should be adjusted as needed based on the monitoring results.

Conclusions

We have the ability to develop options for mitigating climate change in agriculture and food systems. What is nevertheless difficult to achieve is deciding on the combination of options that need to work together in a specific food system to achieve the multiple goals that societies care about and to implement these in a consistent manner that is sensitive to local conditions. For this, we need to acknowledge and work with known power imbalances, vested interests, and fragmented policy making and monitor implementation of outcomes to be able to learn and adjust.

Food system change cannot wait, and neither can action on climate change mitigation, in which the food system has an ever more important part to play.

REFERENCES AND NOTES

1. C. Rosenzweig *et al*., *Nat. Food* **1**, 94–97 (2020).
2. M. Crippa, E. Solazzo, D. Guizzardi, F. N. Tubiello, A. Leip, *Nat. Food* **2**, 198–209 (2021).
3. M. A. Clark *et al*., *Science* **370**, 705–708 (2020).
4. Intergovernmental Panel on Climate Change, "IPCC Sixth Assessment Report: Climate change 2022: Impacts, adaptation, and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change," H.-O. Pörtner *et al*., Eds. (Cambridge Univ. Press, 2022); https://www.ipcc.ch/report/ar6/wg2/.
5. W. Wang, A. Pijl, P. Tarolli, *Nat. Food* **3**, 193–196 (2022).
6. H. E. Froehlich, R. R. Gentry, B. S. Halpern, *Nat. Ecol. Evol.* **2**, 1745–1750 (2018).
7. Food and Agriculture Organization of the United Nations, "The future of food and agriculture: Alternative pathways to 2050" (FAO, 2018); https://www.fao.org/3/I8429EN/i8429en.pdf.
8. M. T. Niles *et al*., *Renew. Agric. Food Syst.* **33**, 297–308 (2018).
9. C. Mbow *et al*., "Food security," In: *IPCC Special Report - Climate Change and Land* (IPCC, 2019); https://www.ipcc.ch/srccl/.
10. World Resources Institute, "Creating a sustainable food future: A menu of solutions to feed nearly 10 billion people by 2050" (WRI, 2018); https://www.wri.org/research/creating-sustainable-food-future.
11. F. N. Tubiello *et al*., *Earth Syst. Sci. Data* **14**, 1795–1809 (2022).
12. A. M. Loboguerrero *et al*., *Sustainability (Basel)* **11**, 1372 (2019).
13. M. Zurek, A. Hebinck, O. Selomane, *Q Open* **1**, qoaa001 (2021).
14. I. D. Brouwer, J. McDermott, R. Ruben, *Glob. Food Sec.* **26**, 100398 (2020).
15. A. Hebinck *et al*., *Glob. Food Secur.* **29**, 100546 (2021).
16. J. Ingram, "Food system models," in *Healthy and Sustainable Food Systems*, M. Lawrence, S. Friel, Eds. (Routledge, 2019); https://doi.org/10.4324/9781351189033.
17. S. J. Vermeulen, B. M. Campbell, J. S. I. Ingram, *Annu. Rev. Environ. Resour.* **37**, 195–222 (2012).
18. J. Poore, T. Nemecek, *Science* **360**, 987–992 (2018).
19. E. Holt-Giménez, A. Shattuck, M. Altieri, H. Herren, S. Gliessman, *J. Sustain. Agric.* **36**, 595–598 (2012).
20. FAO, IFAD, UNICEF, WFP, WHO, "The state of food security and nutrition in the World 2021: Transforming food systems for food security, improved nutrition and affordable healthy diets for all" (FAO, 2021); https://doi.org/10.4060/cb4474en.
21. B. M. Campbell *et al*., *Ecol. Soc.* **22**, art8 (2017).
22. S. Whitfield *et al*., *Nat. Food* **2**, 383–385 (2021).
23. N. D. Mueller *et al*., *Nature* **490**, 254–257 (2012).
24. K. F. Davis *et al*., *Adv. Water Resour.* **99**, 67–75 (2017).
25. Z. Cui *et al*., *Nature* **555**, 363–366 (2018).
26. M. Lessmann, G. H. Ros, M. D. Young, W. de Vries, *Glob. Chang. Biol.* **28**, 1162–1177 (2022).
27. B. W. J. Freeman *et al*., *Glob. Chang. Biol.* **28**, 3795–3811 (2022).
28. S. Leitner *et al*., *Curr. Opin. Environ. Sustain.* **47**, 95–105 (2020).
29. B. Brown, R. Llewellyn, I. Nuberg, *Glob. Food Secur.* **17**, 213–220 (2018).
30. B. M. Roque *et al*., *PLOS ONE* **16**, e0247820 (2021).
31. Y. Hu, H. Cheng, S. Tao, J. L. Schnoor, *Environ. Sci. Technol.* **53**, 12177–12187 (2019).
32. Living with Resistance Project, *Nat. Sustain.* **1**, 632–641 (2018).
33. M. B. Eisen, P. O. Brown, *PLOS Climate* **1**, e0000010 (2022).
34. A. R. Hansen, C. Keenan, G. Sidhu, "Nutritious food foresight: Twelve ways to invest in good food in emerging markets" (Global Knowledge Initiative, 2019); http://globalknowledgeinitiative.org/wp-content/uploads/2019/05/Nutritious-Food-Foresight_GKI-and-GAIN-2019.pdf.
35. T. Searchinger, R. Edwards, D. Mulligan, R. Heimlich, R. Plevin, *Science* **347**, 1420–1422 (2015).
36. M. A. Clark, M. Springmann, J. Hill, D. Tilman, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23357–23362 (2019).
37. N. Wilson, C. L. Cleghorn, L. J. Cobiac, A. Mizdrak, N. Nghiem, *Adv. Nutr.* **10**, S389–S403 (2019).
38. W. Willett *et al*., *Lancet* **393**, 447–492 (2019).
39. K. Hirvonen, Y. Bai, D. Headey, W. A. Masters, *Lancet Glob. Health* **8**, e59–e66 (2020).
40. United Nations Environment Programme, "UNEP food waste index report 2021" (UNEP, 2021); https://www.unep.org/resources/report/unep-food-waste-index-report-2021.
41. E. Wollenberg *et al*., *Glob. Chang. Biol.* **22**, 3859–3864 (2016).
42. I. Stoddard *et al*., *Annu. Rev. Environ. Resour.* **46**, 653–689 (2021).
43. P. Newell, S. Srivastava, L. O. Naess, G. A. Torres Contreras, R. Price, *Wiley Interdiscip. Rev. Clim. Change* **12**, e733 (2021).
44. R. Falkner, *Int. Aff.* **92**, 1107–1125 (2016).
45. C. Okereke, P. Coventry, *Wiley Interdiscip. Rev. Clim. Change* **7**, 834–851 (2016).
46. D. Loorbach, "Transforming climate governance? Why climate governance is failing and what to do about it," in

Transformative Climate Governance, K. Hölscher, N. Frantzeskaki, Eds. (Springer, 2020); pp. 431–445; https://doi.org/10.1007/978-3-030-49040-9_13.
47. J. Fairhead, M. Leach, I. Scoones, *J. Peasant Stud.* **39**, 237–261 (2012).
48. R. Fuchs, C. Brown, M. Rounsevell, *Nature* **586**, 671–673 (2020).
49. W. F. Lamb *et al*., *Global Sustainability* **3**, e17 (2020).
50. A. J. Nightingale *et al*., *Clim. Dev.* **12**, 343–352 (2020).
51. B. K. Sovacool, M. Martiskainen, A. Hook, L. Baker, *Clim. Change* **155**, 581–619 (2019).
52. J. J. Patterson *et al*., *Curr. Opin. Environ. Sustain.* **31**, 1–9 (2018).
53. J. Clapp, *J. Peasant Stud.* **41**, 797–814 (2014).
54. P. Newell, O. Taylor, P. Newell, O. Taylor, *J. Peasant Stud.* **0**, 1–22 (2017).
55. J. Hickel, G. Kallis, *New Polit. Econ.* **25**, 469–486 (2020).
56. A. Khatri-Chhetri *et al*., *Environ. Res. Lett.* **16**, 124044 (2021).
57. M. Hulme, *Zygon* **2015**, 893–905 (2015).
58. J. David Tàbara, J. Jäger, D. Mangalagiu, M. Grasso, *Reg. Environ. Change* **19**, 807–818 (2019).
59. P. Watkiss, M. Benzie, R. J. T. Klein, *Wiley Interdiscip. Rev. Clim. Change* **6**, 541–557 (2015).
60. S. Hasnain, J. S. I. Ingram, M. Zurek, "Mapping the UK Food System - a report for the UKRI Transforming UK Food Systems Programme" (Univ. of Oxford, 2020); https://blogs.exeter.ac.uk/foodsystemimpactscv19/blog/2020/11/26/mapping-the-uk-food-system-a-review/.

ACKNOWLEDGMENTS

Competing interests: The authors declare no competing interests. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 1 April 2022; accepted 25 May 2022
10.1126/science.abo2364

REVIEW

Getting ahead of climate change for ecological adaptation and resilience

Jonathan W. Moore[1]* and Daniel E. Schindler[2]*

Changing the course of Earth's climate is increasingly urgent, but there is also a concurrent need for proactive stewardship of the adaptive capacity of the rapidly changing biosphere. Adaptation ultimately underpins the resilience of Earth's complex systems; species, communities, and ecosystems shift and evolve over time. Yet oncoming changes will seriously challenge current natural resource management and conservation efforts. We review forward-looking conservation approaches to enable adaptation and resilience. Key opportunities include expanding beyond preservationist approaches by including those that enable and facilitate ecological change. Conservation should not just focus on climate change losers but also on proactive management of emerging opportunities. Local efforts to conserve biodiversity and generate habitat complexity will also help to maintain a diversity of future options for an unpredictable future.

The need to rectify human impacts on Earth's climate system is of dire importance, yet there is a parallel urgent need to implement conservation strategies that will enable adaptation of species, promote ecosystem resilience, and maintain the coupled human-ecological systems that support humanity. Even the most aggressive emission reduction strategies will mean that further warming will persist for decades before recovering; it is likely that that warming will exceed 1.5°C for multiple decades, even under very low greenhouse gas (GHG) emissions (*1*). Robust policies and approaches are needed that will provide the best opportunities for promoting ecological adaptation and resilience in a climate-challenged future. Here, we provide an overview of expectations and approaches for maintaining a functioning biosphere that may require thinking outside the box that is typically operated within for managing natural resources and conserving biodiversity.

Ecosystems and society are complex and dynamic systems with connected processes that interact across a hierarchy of scales that can produce resilience—the ability of a system to maintain key functions when disturbed (*2*). System resilience will arise from differential responses, shifts, and turnover in system composition, whether they are genes, populations, species, or habitat conditions. Thus, management for fluidity and connectivity and the expression of heterogeneity within systems can enable resilience (*3*). By contrast, exposure of systems with lower resilience to stressors may push them to undesirable states or compromise their aggregate ecosystem functions, including services to humanity. In the context of climate change, we use the term "adaptation" in a broad sense to refer to these evolutionary, ecological, and social changes that reduce the vulnerability of systems to disturbance (*1*).

"...conservation efforts will need to embrace the dynamic aspects of the biosphere..."

Shift happens

The biosphere is fluid (*2*): Organisms adapt to changing environmental conditions; respond to their predators, competitors, and diseases; move into new habitats; and disappear from places that become inhospitable. These kinds of changes have occurred over the history of the biosphere, including the past 100,000 years (Fig. 1). During the last glacial maximum, 25,000 to 19,000 years ago, the ocean was ~120 m lower, and the world was 10°C colder than it is currently. Ice covered 25% of the terrestrial landscape, and low precipitation and temperatures limited forest cover. As the climate warmed, glaciers melted, and sea level rose rapidly, inundating coasts. Although these global transformations might seem incomprehensible, most species we see today coped with them [for example, most extant vertebrates evolved more than a million years ago (*4*)].

Anthropogenic climate change is driving a new era of rapid change in Earth's climate. The world has already warmed by 1.1°C, and sea level has risen by ~20 cm over the past century (*1*). These changes are small in magnitude relative to historic changes, but there are several critical differences. First, the pace of contemporary change is unusually fast. Second, Earth is predicted to warm to temperatures that have not occurred for more than 100,000 years (*1*). Third, humanity is simultaneously inflicting myriad other stressors that reduce the complexity, fluidity, and resilience of the biosphere. Human-caused extinctions are 100 to 1000 times higher than background rates (*5*), and 38% of terrestrial forest has been converted for human use (*6*). Fourth, humanity has developed infrastructure, governance, cultures, resource management systems, and ethics with only superficial consideration of the oncoming rapid changes to the biosphere that will occur in the near term (*7*).

The biosphere has repeatedly demonstrated its fluidity across scales of biological organization, ranging from the genes within a species to socioecological systems. Intraspecific genetic variation can change across space and time, driven by a dynamic adaptive landscape. For example, contemporary environmental variation and climate extremes, as well as occasional hybridization events, drove rapid evolution in the beaks of Darwin's finches for them to shrink then expand and then shrink again

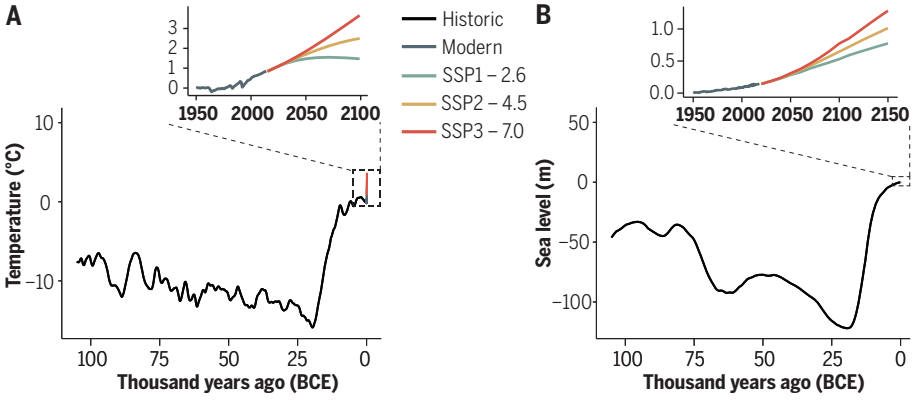


Fig. 1. A dynamic climate, from 100,000 years before present to 2100. Data are anomalies from the present conditions. (**A**) Historically reconstructed surface air temperature (*51*). (Inset) Modern surface air temperature records and projections under three emission scenarios (*1*). (**B**) Reconstructed sea level (*51*). (Inset) Modern sea level records and projections to 2100 (*1*). The projected climates are based on Intergovernmental Panel on Climate Change (IPCC) future GHG emission scenario SSP1-26 (teal), SSP2-45 (yellow), and SSP3-70 (red), representing low, intermediate, and high future emissions, respectively.

[1]Earth to Ocean Research Group, Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada. [2]School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA.
*Corresponding author. Email: jwmoore@sfu.ca (J.W.M.); deschind@uw.edu (D.E.S.)

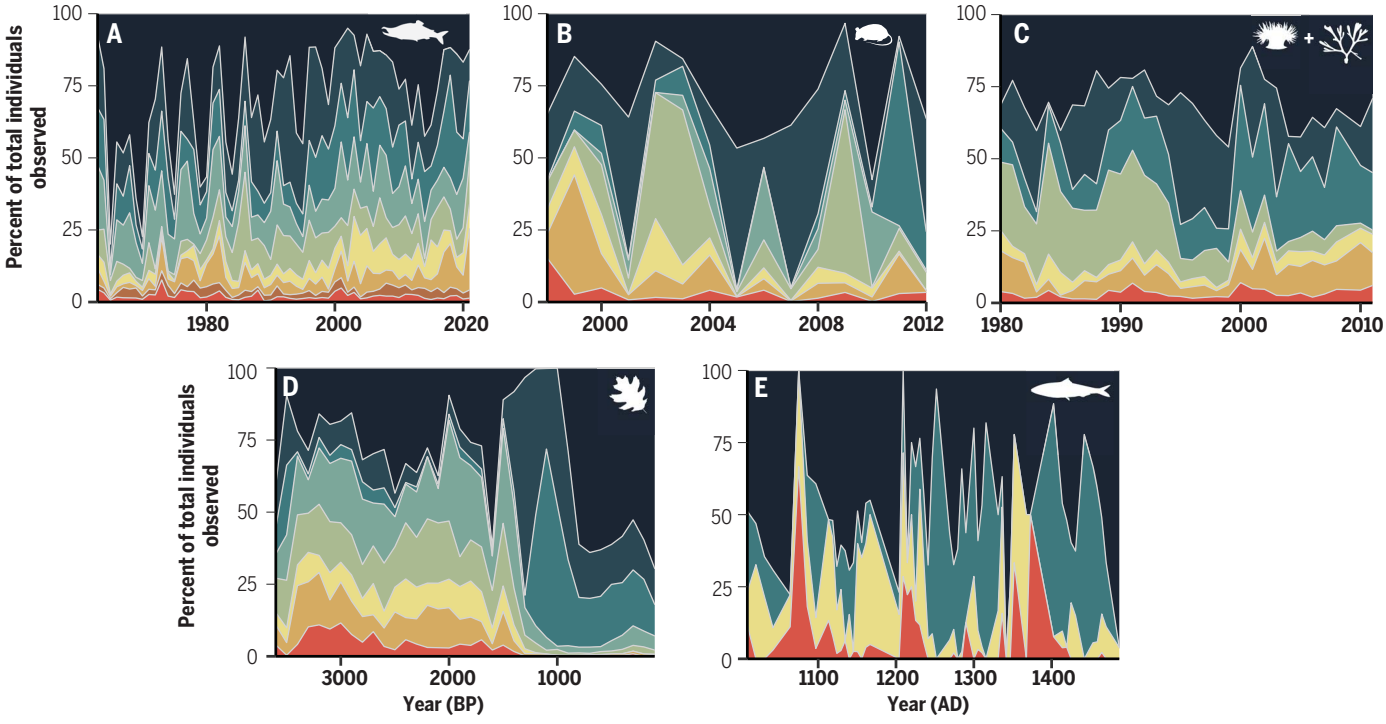


Fig. 2. Shifting biodiversity and turnover within ecosystems. (**A**) Sockeye salmon adult populations returning to nine major rivers in Bristol Bay, Alaska, USA. Data are updated from the Alaska Department of Fish and Game. (**B**) Rodents and marsupials captured in the Atlantic forests of South America (*52*). These data are presented from top (darker cooler color) to bottom (warmest color): *Akodon*, *Oligoryzomys*, *Euryoryzomys*, *Delomys*, *Didelphis*, *Marmosops*, *Marmosa*, and *Nectomys*. (**C**) Kelp forest community structure at San Nicolas Island, CA (*53*). These data include kelp and algae, and invertebrate species: *Cystoseira osmundacea*, *Dictyota binghamiae*, *Diopatra ornata*, *Corallina officinalis*, *Balanophyllia elegans*, *Cucumeria fisheri*, and *Corynactis californica*. (**D**) Fossil pollen counts from Siberia (*54*) for most abundant tree genera: *Pinus*, *Betula*, *Alnus*, *Poaceae*, *Caryophyllaceae*, *Artemisia*, and *Primulaceae*. (**E**) Relative abundance of marine fishes from Santa Barbara basin based on fossil scale counts (*46*): Northern anchovy (*Engraulis mordax*), Pacific hake (*Merluccius productus*), Pacific sardine (*Sardinops sagax*), and Surf perch (*Embiotocidae* spp).

over 30 years (*8*). More generally, driven by both genetic change and plastic life histories, the phenologies of species are shifting in response to contemporary climate and environmental change (*9*).

Species' distributions and range boundaries also continuously change through time. For example, over the past 20,000 years, rapid climate change during glacier retreat drove interlinked processes of adaptation and migration in tree species (*10*, *11*). Pollen records reveal that trees spread poleward rapidly after glacier retreat, aided by small-scale refugia and local adaptations (*10*, *11*).

Contemporary climate change is also redistributing biodiversity, pushing species deeper in oceans, up mountains, and toward the poles (*12*). Terrestrial and marine species are moving poleward an average of 17 and 72 km per decade, respectively (*12*). However, if species' ranges follow shifting thermal regimes, they will not just head poleward given topographic complexities such as mountains, valleys, and coastlines (*13*). Instead, the direction and pace of climate change is extremely variable across the world. There may be climate sinks, such as the tops of mountains or coastal peninsulas, where species are pushed into local dead ends. Alternatively, some areas will be refugia, where environmental conditions remain hospitable. Thus, even simple thermal niche models highlight the complex forcing of climate warming on biological systems across a heterogeneous world.

Ecological communities within a given location also shift unpredictably with climate change as species flourish, wither, or invade (Fig. 2). Pollen records indicate that a plant community in Siberia shifted around 1500 years ago from birch and pine trees to be dominated by herbaceous flowering plants, such as primrose (Fig. 2). On more contemporary time scales, kelp forest communities over 30 years oscillated across years and decades from being primarily brown algae to predominantly cup corals and sea anemones (Fig. 2). Approximately 120,000 years ago, the pelagic fish community in the Humboldt Current off the coast of western South America was composed of small-bodied goby-like fishes when waters were warmer and lower in dissolved oxygen, whereas currently this community is dominated by anchovies that compose upward of 15% of the global annual fish catch (*14*). Analogous dynamics are expressed within metapopulations—for example, the productivity of different salmon populations shift asynchronously over time (Fig. 2). From trees to fishes, biological systems are not characterized by a single static equilibrium state but instead by turnover and compositional shifts across a range of time, space, and taxonomic scales (Fig. 2).

Community changes that involve habitat-forming species, ranging from kelps to corals to trees, alter the fundamental structure of ecosystems. The majority of vegetation communities have gone through large compositional and structural changes over the past 20,000 years (*15*). Contemporary global change is also driving ecosystem transformations, including the production of no-analog ecosystems where new assemblages are emerging with species extinction, invasions, and altered environmental conditions, such as in the transformation of Australian *Eucalyptus* forests to non-native annual grasslands (*16*).

The past 10,000 years of relative climate stability and favorable warm and moist climate conditions led to the emergence of agriculture and modern humanity. Yet even over this period of relative climate stability, people tracked shifting resources over vast distances, diverse cultures and practices adapted to shifting food sources, or in some cases, societies

collapsed during extreme climate events (*1*, *17*). At times during humanity's history, colonization of people to new regions led to overexploitation of species and extirpation (*5*). Human activities have also shaped the biosphere such as by landscape burning, cultivation, and building structures (*17*). Humanity is now in a global crisis of its own making, with human-caused climate change causing risks and harm distributed unevenly around the world.

Forward-looking conservation and management

Natural resource management and conservation efforts will need to embrace the dynamic aspects of the biosphere to help maintain functioning ecosystems and protect biodiversity given oncoming climate change. In Holling and Meffe's visionary 1996 paper (*18*), they described a pathology in natural resource management in which desire to make ecosystems more predictable has led to the development of policies that seek to reduce natural variation in ecological systems. Paradoxically, such approaches can reduce ecosystem resilience to new change that lead to surprises that management agencies are not able to cope with, such as policies of fire suppression leading to catastrophic fires. Here, we adopt their view that "[e]cosystems are moving targets, with multiple potential futures that are uncertain and unpredictable" and emphasize that maintaining the fluid and complex nature of ecosystems is of ever-increasing importance for conservation and resource management (*18*). First, this paradigm would encourage compositional turnover in ecosystems, range shifts, and microevolution in species and flexibility in the human components of ecosystems (Fig. 3). Second, there is concurrent need to also conserve heterogeneity and biodiversity, from genes to species to habitats within ecosystems, that represent evolutionary and ecological options for a changing world, as well as the processes that generate this heterogeneity (Fig. 3). Although there has been extensive attention focused on specific interventions for adaptation to climate change, ranging from assisted gene flow to genome-editing technology such as CRISPR, here instead we focus on approaches to enable the intrinsic resilience of species and places as key components of complex socio-ecological systems during this era of rapid ongoing change.

Species can adapt rapidly to climate change, either through evolutionary adaptation of traits that reduce sensitivity to climate effects, such as physiological thermal tolerance, or decrease their exposure to climate effects, such as by shifting their phenology of key life-history events and geographic ranges (*19*). Oncoming range shifts mandate that conservation science and management grapple with defining what are native, non-native, and invasive species (*20*). Human activities have directly and indirectly increased the movement of many taxa, leading to widespread introduction of species, especially beyond major biogeographic barriers, and corresponding extinctions of native species as well as biotic homogenization. However, species movement into new habitats has always been key to the biosphere's adaptive response to a changing world, and protectionist perspectives could hinder community adaptation (*20*). Perhaps if species are following their projected climate trajectory, then they should be considered "proactively native."

Predicting range shifts by using bioclimatic models to inform spatial conservation efforts are now common. Forward-looking spatial planning for conservation could prioritize refugia, future habitat, corridors for connectivity, and maintenance of landscape heterogeneity. Yet the fine-scale impacts of climate change on species distributions remain highly uncertain because models are used to predict complex systems into no-analog climates (*21*). For example, a study of distributions of European birds found that observed changes varied markedly from predictions for 90% of studied species (*21*). Risk-dispersion approaches are more likely to capture locations that allow species to persist in the future and to include habitats that are underappreciated as critical for certain life stages of target species (*22*). Individuals from populations in nonrefuge areas may provide the genetic material needed for adaptation in the future that could be lost by narrowly focusing on protection of climate refugia alone (*23*). For example, corals from hot regions contain the genes for thermal tolerance that, with sufficient time and connectivity, could enable successful adaption of coral reef systems to warmer waters (*23*). Thus, we caution against the use of overly prescriptive approaches for identifying climate refugia and prioritizing these sites for protection at the expense of maintaining heterogeneity and a range of options across landscapes at local to regional scales.

Most natural resource management is tied explicitly to specific geographic locations. It has also been argued that people's connection to a "sense of place" strengthens such place-based strategies for managing and conserving resources (*24*). However, the composition and identity of places will shift as climate impacts are expressed locally and across the broader landscape. Specific species and ecological communities that are deeply foundational

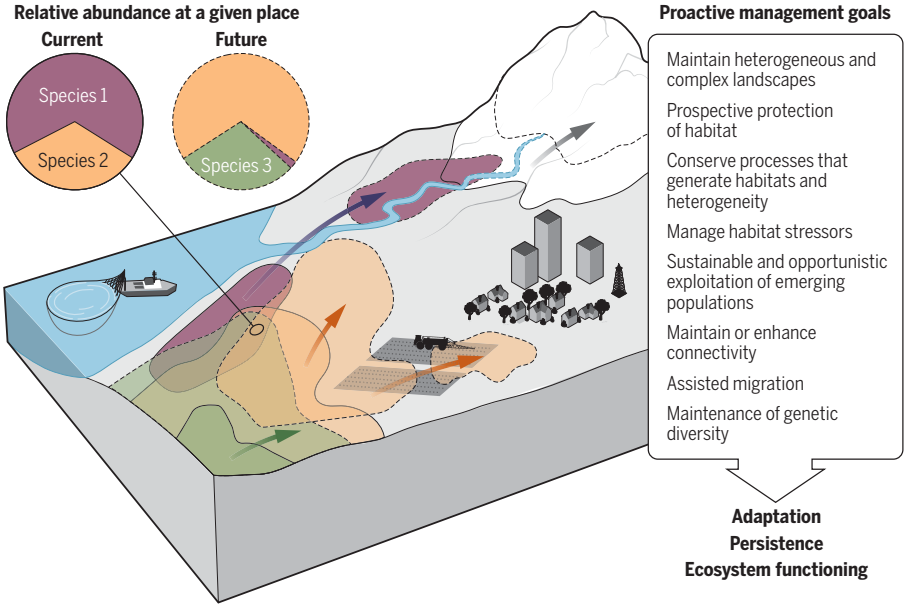


Fig. 3. Expected ecological changes and proactive conservation approaches for enabling adaptation in a rapidly changing world, from genes to ecosystems. Changing climate will transform ecosystems, such as through shifts in snow and ice cover (white region), and also cause changes in the species composition of future ecosystems, including loss of some species and additions of others. Pie charts indicate predicted community composition for current (solid-line chart) and future (dashed-line chart) community composition for a specific location (oval). Ecosystem-based management should enable flexible adaptation of human activities that depend on exploitation of species in specific locations, and the composition of existing protected areas are likely to look fundamentally different from when they were established. Species ranges will change, likely moving poleward and up in elevation as organisms track shifting suitable habitat (solid outlines indicate current ranges, and dashed lines indicate predicted ranges). Maintaining the processes and disturbances that maintain habitat heterogeneity will likely facilitate movement across landscapes and provide options for suitable climate in a climate-altered future.

ILLUSTRATION: KELLIE HOLOSKI/*SCIENCE*

to the identity of places and values of people are already starting to disappear, challenging cultural identities and values (*25*). Place-based management will likely continue to be the norm in most terrestrial and nearshore marine ecosystems, these local management systems will be confronted by unforeseen dynamics that may test their willingness and ability to adapt (*2*).

Reducing cumulative impacts of local stressors can offset the impacts of climate change in some situations. For example, thermally sensitive freshwater fishes are threatened by warming water temperatures in some regions. However, river water temperatures are not just affected by climate but also by removal of riparian forests, an extensive impact in many watersheds attributable to logging, agriculture, or urbanization. One study of a river in western North America estimated that intact riparian forests can reduce maximum summer water temperatures by 7°C, offsetting even a 4°C increase in future air temperatures by 2080 (*26*). Similarly, the combined impacts of climate warming and nonpoint nutrient pollution are increasing the eutrophication of coastal and inland waters and associated hypoxia, toxic algae blooms, and fish kills (*27*). Thus, cumulative stressors can exacerbate many impacts of climate change. Reciprocally, there is enormous opportunity for local habitat protection and restoration to offset or reduce the local consequences of climate change (*16*, *28*).

A primary goal for conservation policy to maintain resilience to global change is to conserve heterogeneity within ecosystems and their biota (*28*). Habitat heterogeneity is expressed across a range of spatial and temporal scales and is produced through a diverse assortment of biotic and abiotic processes. Restoration and maintenance of this heterogeneity has traditionally focused on the features of habitat themselves, rather than on the processes that generate them in the first place (*18*). Emerging restoration paradigms emphasize the latter—that conserving the processes that generate habitat heterogeneity is more likely to produce the features important for ecosystem resilience and species persistence (*29*). Long-term efforts to suppress and control disturbances have likely eroded the resilience of ecosystems and their ability to adapt to the new conditions that lay ahead with climate change and further growth of humanity (*18*). Protection of complex landscapes as dynamic habitat mosaics will enable their composition to turn over and maintain function. For example, salmon fisheries in Alaska rivers are stabilized within heterogeneous watersheds because of the complementary dynamics among tributary populations as they respond differentially to local-scale expressions of regional climate forcing (*30*).

A primary approach in conservation has been to establish permanent protected areas that are intended to shield habitats and species from cumulative human pressures. However, many species may move away from static protected areas, and the composition of protected habitats will turn over as climate imposes new conditions on the landscape (*31*). Thus, there is a need to expect that this approach may fail to conserve the habitats and species that originally motivated the formation of protected areas and that new collections of species may emerge (*31*). Recent emphasis on networks of protected areas with ecological connectivity among individual sites will enable the dispersal needed for species adaptation and persistence and potential for community turnover needed for broad-scale adaptation to new climate conditions (*22*, *23*, *32*). Beyond protected areas, the effective management of "working lands" as ever-evolving socio-ecological systems will also be a key component for proactive maintenance of ecosystem services, connectivity, and local biodiversity at landscape scales (*16*, *33*).

Dynamic management approaches have recently emerged to more effectively protect species in a state of flux. In such dynamic management approaches, spatial data on environmental conditions, animal movements, and the distribution of resource users (such as fishers) are integrated to adaptively shift protected areas to achieve conservation goals but also allow other sectors to use the resource. In marine ecosystems, dynamic ocean management has become an increasingly popular idea for protecting mobile species while enabling resource extraction activities that do not cause long-term ecological damage (*34*). The principal strategy is to provide moving spatial protection to mobile vulnerable species by continuously updating where they are susceptible to human activities, such as by-catch in fisheries for more common target species. Such dynamic approaches in ocean conservation offer promise but may be challenged by limited enforcement; cross-jurisdictional movements of species (*35*); and the need for coordinated and intensive monitoring, assessment, and forecasting of the status and distribution of vulnerable and targeted species. Dynamic management approaches also hold promise in some terrestrial systems, such as renting wetlands from farmers during specific seasons for migratory waterbirds (*36*). However, human impacts to landscapes can be long lasting, and thus dynamic approaches in terrestrial systems may be more spatially limited or responsive over longer time scales (*37*).

Management can also protect or restore habitats that are important future options for species. Human activities often eliminate viable future options for species, whether through invasive species, migration barriers, or habitat degradation. For example, coastal development often blocks the upslope migration of coastal wetlands in response to sea-level rise (*38*). Amphibians use both montane wetlands and lakes for breeding habitats. Yet even as climate change increasingly dries up wetlands, the more climate-resilient lakes have largely been removed as viable options for amphibians by the systematic introduction of predatory non-native trout to historically fishless lakes (*39*). Forward-looking protection or restoration actions can provide the future habitats that species might need to survive in a climate-altered future.

Oncoming ecosystem transformations from climate change are creating opportunities for some species but also frontiers for extractive industries, calling for preemptive habitat protection of nascent ecosystems. Ice retreat, whether it is in the polar oceans or glaciers, is

Box 1. Ecosystem transformation of the Arctic Ocean.

In the Arctic, what used to be covered in ice throughout the year is now increasingly open water during the summer. Minimum sea ice extent decreased by 30% over the past 29 years (*1*). Some species are threatened by ice retreat, with polar bears being the poster-child climate-change loser. Ice retreat is altering oceanic processes in complex ways that are difficult to predict; for example, the retreat of ice in the northern Bering Sea has changed food webs such as through reductions in ice-reliant and lipid-rich algae. Yet general predictions are for many fish stocks to shift poleward and for a >75% increase in biomass by 2100 in some Arctic marine regions (*40*).

The Arctic transformation to ice-free summers is bringing human industrial pressures. With the potential of increased fish stocks and increased access, there is promise of emerging fisheries. New shipping corridors are also opening up, driving rapid increases in shipping traffic and associated environmental risks, such as ship strikes to marine mammals (*55*). The continental shelf in the Arctic Ocean is also thought to contain substantial fossil fuel reserves—31% of unexplored natural gas and 13% of oil.

There are emerging efforts for forward-looking and precautionary governance of this rapidly transforming Arctic and its oncoming pressures. Canada and the United States have a temporary moratorium on Arctic offshore oil and gas licensing. Recent international fishing agreements emphasize precaution (*56*). Yet governance of this climate frontier remains complex and fraught among the five bordering states and other interested parties.

creating new ecosystems. For example, the Arctic Ocean is rapidly transitioning to having ice-free summers. Although this transformation is threatening ice-dependent species, ranging from polar bears to diatoms, major increases in Arctic fish production are expected over the next century (*40*), creating opportunities for fisheries. However, ice-free summers are also increasing Arctic shipping traffic as well as oil and gas exploration, which pose major environmental risks. Emerging initiatives are working toward forward-looking management of this multijurisdictional and contested resource frontier (Box 1). Alpine Glacier retreat is raising similar challenges. Glacier loss threatens water supply for hundreds of millions of people, as well as ecological flows and temperatures in downstream aquatic ecosystems (*1*). Yet in some local regions in western North America, glacier retreat is exposing new habitats, such as for migratory Pacific salmon that support fisheries. In the next 80 years under a moderate climate scenario, glacier retreat is predicted to create ~6000 km of new streams accessible for Pacific salmon and other stream-associated species (*41*). However, mineral development is targeting the retreating edges of icefields to mine newly exposed deposits (*42*). Thus, although climate change is creating local opportunities and options for species, extractive industries that can cause long-lasting environmental damage could preemptively undermine these options. Proactive protection of these frontier habitats would enable the realization of these future opportunities for species and associated sustainable use.

Industries and communities dependent on natural resources could more effectively realize the benefits of taxonomic turnover for stabilizing ecosystem services. For example, fishing communities are more resilient to climate perturbations when fisheries are allowed to harvest new species to replace those that become locally extirpated because of climate fluctuations (*43*, *44*). Flexibility in what fishers can capture provides for stability, particularly given the tendency of fish populations to boom and bust in their population dynamics (*45*, *46*). For example, after a climate regime shift in the North Pacific Ocean in 1989, Pacific herring (*Clupea pallasii*) and salmon (*Oncorhynchus* spp.) in the northern Bering Sea declined, whereas walleye pollock (*Gadus chalcogrammus*) in coastal Southeast Alaska increased; communities that had the required permits and gear could switch to catching walleye pollock and thus suffered little economic consequence of the regime shift (*43*). Similarly, fishers in the northeast United States switched to catching more fluke (*Paralichthys dentatus*) as hake (*Merluccius bilinearis*) became less abundant, buffering their revenues despite climate-driven changes to the abundance and distribution of these fishes (*44*). Given that climate change will shift the distribution of resources across space and time, legitimate policy options might allow exploiters to track transient harvest opportunities as they become available (*47*), instead of treating such opportunistic behavior as roving banditry (*24*). Like the dynamic habitat protection strategies described above, such harvesting strategies will require rapid assessment of emerging resources and a high level of control of exploiters to relieve pressure on species that become rare or unproductive.

Environmental decision-making and resource management strategies have a history of putting little emphasis on the future, and oncoming climate change necessitates a shift in this perspective. Myopic approaches to resource management often derive from the application of economic models that make strict assumptions through social discounting about the value of ecosystems, or the costs of restoring them, in the future. But future generations will bear nearly all of the brunt of ongoing climate change, even if we can rapidly reduce GHG emissions (*48*). When assessing the costs of protecting habitat now, versus the potential costs of damage to ecosystems that will be encountered by the generations in a climate-altered future, a lower discount rate is reasonable, particularly given the uncertainties of how the future of ecosystems will unfold (*49*). Given that different genes, populations, species, and habitats all provide options for an uncertain future, there is an increasingly urgent need to stem the rate of loss of these potential options. The approaches we have highlighted here may enable species and ecosystems to function and persist in a warmer future. These approaches place emphasis on future options for species and ecosystems rather than simply treating them as a commodity of the present.

Conclusions

Developing and implementing successful conservation science for a climate-altered future will require proactive application of existing and emerging approaches. The futures considered by humanity must be assumed to be different from what were experienced in the past, and these futures are deeply uncertain (*7*, *50*). Expectations should be for social, ecological, and evolutionary change, and science and policies should develop strategies for enabling change, monitoring it, and balancing the risks and opportunities it presents to the biosphere and humanity. Conservation approaches should facilitate evolution, dispersal, and compositional turnover in ecosystems. A fundamental stumbling block will be to reconcile such approaches that maintain community turnover while preventing extirpations of biodiversity that provide future options. Although scientific approaches may illuminate tradeoffs in achieving each of these potentially conflicting goals under different conservation strategies, deep engagement with a diversity of stakeholders and the values they hold should ultimately point the way forward. In some cases, oncoming changes as well as proactive approaches will challenge common perspectives and values that were built on a baseline that was assumed to be stationary (*25*). Maintaining and restoring the processes that generate heterogeneity in habitats, genes, and communities should be prioritized for maintaining ecological options for the future. Thus, most of the conservation strategies needed to enable adaptation in species and ecosystems and to minimize climate impacts will play out at local to regional scales, even as reducing emissions of GHG to allow recovery of the climate system will require prompt global coordination. Climate impacts may completely overwhelm local actions, and actions may have unintended consequences that demand further adaptation. Regardless, even as humanity strives to reduce GHG emissions, there is urgent need to protect ecological options and enable adaptation of species and ecosystems to a shifting and unpredictable Anthropocene.

REFERENCES AND NOTES

1. H.-O. Portner *et al.*, *IPCC, 2022: Climate Change 2022: Impacts, Adaptations, and Vulnerability* (IPCC, 2022).
2. L. H. Gunderson, C. S. Holling, *Panarchy: Understanding Transformations in Human and Natural Systems* (Island Press, 2002).
3. R. Biggs *et al.*, *Annu. Rev. Environ. Resour.* **37**, 421–448 (2012).
4. J. T. Weir, D. Schluter, *Science* **315**, 1574–1576 (2007).
5. S. L. Pimm, G. J. Russell, J. L. Gittleman, T. M. Brooks, *Science* **269**, 347–350 (1995).
6. W. Steffen *et al.*, *Science* **347**, 1259855 (2015).
7. P. Kareiva, E. Fuller, *Glob. Policy* **7**, 107–118 (2016).
8. P. R. Grant, B. R. Grant, *Science* **296**, 707–711 (2002).
9. C. Parmesan, G. Yohe, *Nature* **421**, 37–42 (2003).
10. M. B. Davis, R. G. Shaw, *Science* **292**, 673–679 (2001).
11. R. G. Pearson, *Trends Ecol. Evol.* **21**, 111–113 (2006).
12. G. T. Pecl *et al.*, *Science* **355**, eaai9214 (2017).
13. M. T. Burrows *et al.*, *Nature* **507**, 492–495 (2014).
14. R. Salvatteci *et al.*, *Science* **375**, 101–104 (2022).
15. C. Nolan *et al.*, *Science* **361**, 920–923 (2018).
16. R. J. Hobbs *et al.*, *Front. Ecol. Environ.* **12**, 557–564 (2014).
17. E. C. Ellis, *Annu. Rev. Environ. Resour.* **46**, 1–33 (2021).
18. C. S. Holling, G. K. Meffe, *Conserv. Biol.* **10**, 328–337 (1996).
19. A. A. Hoffmann, C. M. Sgrò, *Nature* **470**, 479–485 (2011).
20. G. R. Walther *et al.*, *Trends Ecol. Evol.* **24**, 686–693 (2009).
21. M. B. Araújo, C. Rahbek, *Science* **313**, 1396–1397 (2006).
22. H. L. Beyer *et al.*, *Conserv. Lett.* **11**, 1–10 (2018).
23. T. E. Walsworth *et al.*, *Nat. Clim. Chang.* **9**, 632–636 (2019).
24. F. Berkes *et al.*, *Science* **311**, 1557–1558 (2006).
25. W. N. Adger, J. Barnett, K. Brown, N. Marshall, K. O'Brien, *Nat. Clim. Chang.* **3**, 112–117 (2013).
26. S. M. Wondzell, M. Diabat, R. Haggerty, *J. Am. Water Resour. Assoc.* **55**, 116–132 (2019).
27. S. R. Carpenter *et al.*, *Ecolog. Appl.* **8**, 559–568 (1998).
28. J. J. Lawler *et al.*, *Front. Ecol. Environ.* **8**, 35–43 (2010).
29. T. J. Beechie *et al.*, *Bioscience* **60**, 209–222 (2010).
30. R. Hilborn, T. P. Quinn, D. E. Schindler, D. E. Rogers, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6564–6568 (2003).
31. P. R. Elsen, W. B. Monahan, E. R. Dougherty, A. M. Merenlender, *Sci. Adv.* **6**, eaay0814 (2020).
32. C. A. Correa Ayram, M. E. Mendoza, A. Etter, D. R. P. Salicrup, *Prog. Phys. Geogr.* **40**, 7–37 (2016).

33. C. Kremen, A. M. Merenlender, *Science* **362**, eaau6020 (2018).
34. M. Pons *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2114508119 (2022).
35. M. L. Pinsky *et al*., *Science* **360**, 1189–1191 (2018).
36. M. D. Reynolds *et al*., *Sci. Adv.* **3**, e1700707 (2017).
37. W. K. Oestreich, M. S. Chapman, L. B. Crowder, *Front. Ecol. Environ.* **18**, 496–504 (2020).
38. K. Thorne *et al*., *Sci. Adv.* **4**, eaao3270 (2018).
39. M. E. Ryan, W. J. Palen, M. J. Adams, R. M. Rochefort, *Front. Ecol. Environ.* **12**, 232–240 (2014).
40. D. G. Boyce, H. K. Lotze, D. P. Tittensor, D. A. Carozza, B. Worm, *Nat. Commun.* **11**, 2235 (2020).
41. K. J. Pitman *et al*., *Nat. Commun.* **12**, 6816 (2021).
42. W. Colgan, H. Højmark Thomsen, M. Citterio, *Geol. Surv. Denmark Greenl. Bull.* **33**, 61–64 (2015).
43. T. J. Cline, D. E. Schindler, R. Hilborn, *Nat. Commun.* **8**, 14042 (2017).
44. E. A. Papaioannou *et al*., *Front. Mar. Sci.* **8**, 1–25 (2021).
45. K. A. Vert-pre, R. O. Amoroso, O. P. Jensen, R. Hilborn, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1779–1784 (2013).
46. S. McClatchie, I. L. Hendy, A. R. Thompson, W. Watson, *Geophys. Res. Lett.* **44**, 1877–1885 (2017).
47. J. E. Cinner *et al*., *Nat. Clim. Chang.* **8**, 117–123 (2018).
48. W. Thiery *et al*., *Science* **374**, 158–160 (2021).
49. S. Lewandowsky, M. C. Freeman, M. E. Mann, *Global Planet. Change* **156**, 155–166 (2017).
50. D. E. Schindler, R. Hilborn, *Science* **347**, 953–954 (2015).
51. R. Bintanja, R. S. W. van de Wal, J. Oerlemans, *Nature* **437**, 125–128 (2005).
52. R. S. Bovendorp *et al*., *Ecology* **98**, 2226 (2017).
53. M. C. Kenner *et al*., *Ecology* **94**, 2654–2654 (2013).
54. X. Cao *et al*., *Earth Syst. Sci. Data* **12**, 119–135 (2020).
55. D. D. W. Hauser, K. L. Laidre, H. L. Stern, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7617–7622 (2018).
56. A. N. Vylegzhanin, O. R. Young, P. A. Berkman, *Mar. Policy* **118**, 104001 (2020).

ACKNOWLEDGMENTS

S. Wilson and D. Scurfield assisted with figure preparation. We thank the Gordon and Betty Moore Foundation for support and S. McClatchie and I. Hendy for sharing data. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 4 April 2022; accepted 19 May 2022
10.1126/science.abo3608

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

Giant tortoises, like Diego here, a centenarian in the Galápagos Islands, can delay senescence to become exceptionally long lived, particularly when housed in protected reserves.

PHOTO: RPBMEDIA/ ISTOCKPHOTO

AGING

How to cheat senescence?

Compared with birds and mammals, herpetiles, especially turtles and tortoises, are well-known examples of extremely long-lived animals that show little evidence of age-related decline (see the Perspective by Austad and Finch). By comparing aging rates and longevity across 77 species of reptiles and amphibians, Reinke *et al.* found considerable variation in senescence and elucidated some of the drivers of these differences in nature. In another paper, Da Silva *et al.* studied turtles and tortoises in zoos and found clear evidence that negligible senescence occurs under controlled conditions. —SNV

Science, abm0151, abl7811, this issue p. 1459, this issue p. 1466;see also adc9442, p. 1384

MICROBIOLOGY

A magnificent megabacterium

We usually think of bacteria as microscopic isolated cells or colonies. Sampling a mangrove swamp, Volland *et al*. found an unusually large, sulfur-oxidizing bacterium with a complex membrane organization and predicted life cycle (see the Perspective by Levin). Using a range of microscopy techniques, the authors observed highly polyploid cells with DNA and ribosomes compartmentalized within membranes. Single cells of the bacterium, dubbed *Candidatus* Thiomargarita magnifica, although thin and tubular, stretched more than a centimeter in length. —MAF

Science, abb3634, this issue p. 1453; see also adc9387, p. 1379

NEURODEVELOPMENT

Origins of astrocyte populations

In the developing brain, radial glia are the stem cells that generate astrocytes. By mid-gestation in humans, radial glia form two populations derived from the pial surface of the brain or the ventricular surface. Allen *et al*. used lineage tracing and single-cell transcriptomics in the developing human brain to understand how these two populations of radial glia relate to astrocyte diversity. Astrocytes that populate the brain's gray matter, which is densely packed with neuronal cell bodies, are generated by stem cells in the ventricular zone. Stem cells in the outer subventricular zone produce a different suite of astrocytes restricted to the brain's white matter. —PJH

Science, abm5224, this issue p. 1441

OPIOID USE DISORDER

Reducing withdrawal symptoms

In patients with opioid use disorder, opioid withdrawal is associated with long-term sleep disturbances that can affect drug-seeking behavior and overall treatment outcome. Previous studies have shown that orexin (hypocretin) signaling is implicated in sleep continuity, and its inhibition could reduce withdrawal symptoms in rodents. Huhn *et al.* evaluated the effects of the orexin receptor antagonist suvorexant in 38 subjects with opioid use disorder while they were undergoing buprenorphine treatment as part of opioid withdrawal therapy. Suvorexant improved sleep compared with placebo, with no evidence of abuse potential, suggesting that targeting orexin signaling might

reduce the negative effects of opioid withdrawal on sleep quality. —MM

Sci. Transl. Med. **14**, eabn8238 (2022).

NETWORK SCIENCE

A cost-benefit approach to criticality

The flow of information through networks is maximized near a critical point at which small changes to a system result in large changes in its function. However, every state has its drawbacks, and for criticality, there is a trade-off between sensitivity and stability. Poel *et al*. studied startle cascades in schools of golden shiner fish through the lens of costs of criticality. Schools fine-tuned their distance from criticality primarily by school density in order to match perceived risk. In relatively safe or high-predation environments, the costs of errant startles outweighed the benefits of fast responses and the school remained subcritical. Only when predation risk was difficult to distinguish was the payoff of criticality worthwhile for individual fish. —JSH

*Sci. Adv.*10.1126/sciadv.abm6385 (2022).

GEOMORPHOLOGY

Sediment traps and drains

Humans dramatically change the amount of sediment that makes it to the oceans and seas by damming rivers or through land-use changes. Dethier *et al*. used satellite imagery from the mid-1980s onward, ground truthed with over 100,000 measurements, to estimate the sediment flux from 414 rivers worldwide (see the Perspective by Zarfl and Dunn). Dams have substantially reduced sediment flux in the global north, whereas land-use changes have increased sediment flux in the global south. These observations can help to guide policy decisions regarding critical water resources. —BG

Science, abn7980, this issue p. 1447; see also abq6986, p. 1385

MARINE LIPIDS

Homeoviscous adaptation in plankton

The lipid composition of organisms will change as surrounding physicochemical conditions change to ensure that membranes and other essential biological functions stay intact, a process known as homeoviscous adaptation. Holm *et al*. used high-resolution mass spectrometry to analyze the lipid composition of surface layer plankton sampled largely from the Atlantic Ocean (see the Perspective by Sepúlveda and Cantarero). Although a few lipid species constitute the bulk of the plankton lipidome, organisms living in cold (−2°C) regions have threefold more unsaturated fatty acids than those in warm (29°C) waters. Declines in unsaturation of planktonic essential fatty acids under different scenarios of climate change could have negative consequences for food webs and ultimately for fisheries. —CA

Science, abn7455, this issue p. 1487; see also abo5235, p. 1378

FOREST ECOLOGY

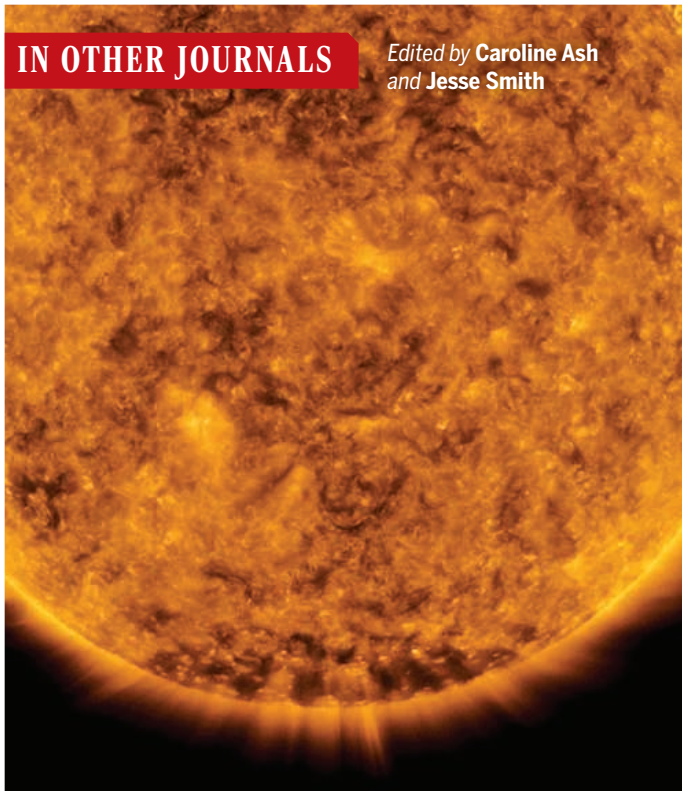
A preindustrial forest carbon sink

Past changes in forest biomass can help to predict how forests—and their ability to sequester carbon—will change in the future. Raiho *et al*. modeled changes in biomass in the Midwestern United States over the past 10,000 years using data from preindustrial forest surveys combined with fossil pollen records. They found that forest biomass was not stable before industrialization, as many studies have assumed. Instead, biomass doubled over the past 8000 years due to increasing forest area after glaciers receded and to changing tree species composition. The biomass accumulated over millennia has been lost in only 150 years of deforestation. —BEL

Science, abk3126, this issue p. 1491

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**



STELLAR ASTROPHYSICS

A higher metallicity for the Sun

The abundances of chemical elements in the Sun are used as a standard composition against which other stars, planets, and galaxies are compared. Precisely determining the solar composition is complex, requiring analysis of thousands of spectral lines, incorporation of uncertain atomic and molecular data, and construction of a self-consistent radiative transfer model of the Sun's atmosphere. Magg *et al.* used this approach to determine updated abundances for eight of the most common elements. They found a higher fraction of elements heavier than helium (known as the metallicity) than was found with previous methods. The results could resolve a discrepancy between the solar composition and constraints from helioseismology. —KTS *Astron. Astrophys.* **661**, A140 (2022).

The Sun has a higher abundance of elements heavier than helium than was previously thought.

HUMAN GENOMICS

A close-up of coronary arteries

Coronary artery disease involves a narrowing of the blood vessels that supply the heart muscle. It is caused by a complex interplay of injury to the blood vessel walls and plaque buildup, and despite decades of medical advances, remains the leading cause of death worldwide. To understand the underlying disease mechanisms, Turner *et al.* performed single-nucleus analysis of chromatin accessibility in tissue samples obtained from 41 patients with coronary artery disease. By combining the resulting single-nucleus atlas with genetic variants implicated in risk for coronary artery

PHOTO: NASA/SOLAR DYNAMICS OBSERVATORY/JOY NG

CONSERVATION

People count bumblebees

A great many species are declining in numbers. Truly understanding the specifics of such declines across species requires large amounts of data that are not trivial to collect. Community science allows nonprofessionals to contribute data that are otherwise difficult to amass. Whitehorn *et al.* used data collected by community scientists during "bee walks," monthly species identification surveys conducted across the United Kingdom over 10 years. The data have revealed associations among habitat, climate, and bumble bee species presence and abundance. Although associations between these factors varied widely, wetland and riparian habitats were consistently associated with higher values. These results highlight the need to understand specific details about each species when planning conservation actions, but also show that protecting highly threatened wetland and riparian habitats should be a high priority for British bumble bee species. —SNV *J. App. Ecol.* 10.1111/1365-2664.14191 (2022).

The buff-tailed bumble bee, *Bombus terrestris*, is a widespread and abundant species in the UK.

PHOTO: IAN600F/ISTOCKPHOTO

disease, the authors were able to identify some of the biological mechanisms underpinning these risk variants, which may help to pave the way for future therapeutic interventions. —YN

Nat. Genet. **54**, 804 (2022).

AUTOIMMUNITY

TLR7 mutation drives human lupus

Toll-like receptors (TLRs) are a family of cell surface sensors that can detect pathogens and mediate inflammatory responses. TLR7 is known for its role in detecting single-stranded RNA and has been associated with auto-antibody production. Brown *et al.* discovered gain-of-function mutations in the *TLR7* gene in several individuals with the autoimmune disease systemic lupus erythematosus (SLE). When engineered into mice, the abnormal *TLR7* variant made immune cells more sensitive to the molecule guanosine, increased the survival of activated B cells, and caused tissue damage similar to that seen in human SLE. Autoimmune features could be prevented in mice by inhibiting production of the MYD88 protein, which TLR7 normally uses to activate B cells. These findings indicate that the TLR7–MYD88 pathway may be targetable for treating patients with lupus. —PNK

Nature **605**, 349 (2022).

NANOSENSORS

Deep detection of fluorescence signals

Wide absorption fluorescent nanoparticles are valuable probes for chemical reactions, drug efficacy, and disease detection, but their use is largely limited to in vitro or shallow in vivo studies because signals can overlap with autofluorescence from target tissues. Using nanosensors that retain a constant absorption but have changing fluorescence in the presence of an analyte, Koman *et al.* showed that by oscillating the excitation around the absorption peak, they could filter out any constant background signal. This improved the sensitivity by more than 50-fold in a range of tissues and phantoms, making it possible to track responses to a range of molecules at depths beyond 5 centimeters or to track a chemotherapeutic transcranially at a depth of 2.4 centimeters in a porcine brain. —MSL

Nat. Nanotechnol. 10.1038/s41565-022-01136-x (2022).

MICROBIOME

Urban gut spillover

As urbanization spreads, other species have taken advantage of the alternative food sources that cities can supply. As a result, some human-wildlife contact can be very intimate indeed. Dillard *et al.* have discovered that urban species as diverse as *Anolis* lizards, coyotes, and white-crowned sparrows have acquired gut bacterial lineages associated with urban humans. In particular, *Anolis* lizards have picked up a *Bacteroides* sequence variant that is overrepresented among people and associated with a diet high in animal fat and protein. A major question is what effect might loss of an indigenous microbiota have on the urban host species' immune responsiveness, growth, and fitness? —CA

eLife **11**, 76381 (2022).

EDUCATION

Switching mentors as a graduate student

Switching mentors during graduate school is often done quietly. Now, with new emphasis on student well-being during graduate training, it is time to acknowledge and explore why and how switching happens. Schmidt and Hansson engaged 19 graduate students in semi-structured interviews to learn more about their reasons for switching mentors. Students described the long process as stressful, but they also reported feeling renewed and energized once the switch was complete. Although students described the advantages of making a switch, they also reported fearing consequences that could affect them long after graduation. Data also highlighted the importance of program directors holding uncomfortable discussions and advocating for the student, and of academic institutions establishing clear guidelines that are documented and implemented in practice. —MMC

Studies in Graduate and Postdoctoral Education **13**, 54 (2022).

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

ORGANIC CHEMISTRY

Sizing rings with carbon–hydrogen oxidation

Catalytic reactions targeting carbon–hydrogen bonds have proliferated in recent years, but selecting a particular site in a hydrocarbon chain remains a great challenge. Chan *et al*. report that precise optimization of the nitrogen ligands coordinating a palladium catalyst can tune the ring size accessed in lactone formation. More specifically, with reagents comprising a hydrocarbon chain capped at each end by an acid, the choice of ligand determines whether oxidative ring closure will occur two or three carbons away from an end group. —JSY

Science, abq3048, this issue p. 1481

NEUROSCIENCE

Adjusting GABA signaling with PKA

In neurons, the adhesion molecule neuroligin-2 (NL2) promotes the surface localization of receptors for the inhibitory neurotransmitter GABA, a role that is inhibited by the phosphorylation of NL2. Halff *et al*. found that NL2 phosphorylation was mediated by the kinase PKA. In cultured rat neurons, this phosphorylation event induced the dispersal of NL2 from the synaptic surface, which decreased the surface abundance of the GABA receptor and dampened inhibitory currents. In contrast, GABAergic signaling was enhanced in neurons expressing a form of NL2 that could not be phosphorylated. —LKF

Sci. Signal. **15**, eabg2505 (2022).

MUCOSAL IMMUNOLOGY

Organized immunity in the oral mucosa

The oral cavity is the first site of contact with foodborne pathogens, yet how adaptive immune responses to antigens in the oral mucosa are regulated is not well understood. Barreto de Albuquerque *et al*. modeled foodborne listeriosis using oral *Listeria monocytogenes* (Lm) infection in mice. Lm drained from the oral mucosa to mandibular lymph nodes (mandLNs), resulting in local $CD8^+$ T cell activation and effector (TEFF) generation. mandLN-primed TEFF disseminated to the lung, oral mucosa, and secondary lymph nodes to mount protective effector responses. The cells did not disseminate to the small intestine because of reduced expression levels of gut-homing receptors. Thus, priming of $CD8^+$ T cells in mandLNs contributes to host protection, extending the concept of compartmentalized immune responses within the gastrointestinal tract. —HMI

Sci. Immunol. **7**, eabf1861 (2022).

SOCIAL SCIENCES

Conversations temporarily reduce polarization

Compared with people who did not know that they were talking to a cross-partisan stranger, when people aligned with one party knew that they were talking with someone from an opposing party, they reported feeling better toward their partner. Santoro and Broockman recruited participants to have videochat conversations about their perfect day, and randomly chose some pairs to receive information about the partisan preferences of their partner. Although the reduction in feelings of polarization was large, it was temporary, with the effects disappearing after 3 months. The promise of such personal interactions for support of democratic norms remains unclear, because participants showed little change in their attitudes toward democratic accountability. —RT

Sci. Adv. 10.1126/sciadv.abn5515 (2022).

ORGANIC CHEMISTRY

Making complex amines from nitriles

Reducing the carbon–nitrogen triple bond in nitriles is a common, convenient way to make primary amines for a vast array of chemical applications. Chandrashekhar *et al*. now report a triphosphine nickel catalyst that can couple nitriles with preexisting amines under hydrogen to produce more complex secondary (disubstituted) and tertiary (trisubstituted) amines as well (see the Perspective by Monguchi). The relative rates of the hydrogen addition and nitrogen exchange steps are key to the selective generation of a single product. —JSY

Science, abn7565, this issue p. 1433; see also abq7782, p. 1382

DNA DAMAGE REPAIR

Teaching an enzyme a new trick

Small molecules that activate enzymes commonly operate using allosteric control. Michel *et al*. describe the mode of action of a small-molecule activator that binds to the active site of 8-oxo guanine DNA glycosylase 1 (OGG1) and catalyzes a biochemical reaction not found in the naïve protein. The molecule enables the protein to completely cleave the damaged DNA strand, which results in an overall increased repair of oxidative DNA damage. The underlying creation of a new repair pathway has potential medical application in diseases related to oxidative DNA damage. —DJ

Science, abf8980, this issue p. 1471

CRISPR

Smaller, simpler, and stealthier

The evolutionary origin of CRISPR-Cas9 has been traced to the transposon-encoded nuclease IscB. A fraction of the size of Cas9, IscB is just as active as Cas9 in RNA-guided DNA cleavage in test tubes. How did it gradually morph into Cas9? Can we take advantage of its miniature size for expanded genome-editing usage? By solving cryo–electron microscopy structures of IscB-ωRNA, Schuler *et al*. provide a high-resolution explanation for IscB's structural and mechanistic similarities to and differences from Cas9. Structure-inspired ideas allowed for the engineering of stealthier versions of IscB and further improvement of IscB's genome-editing activity in human cells. —DJ

Science, abq7220, this issue p. 1476

IMMUNOMETABOLISM

RETooling the anti-TB arsenal

Although tumor necrosis factor (TNF) acts an important tuberculosis (TB) resistance factor, too much of this cytokine can kill infected macrophages and allow mycobacteria to escape and grow uncontrolled in the extracellular space. Using a zebrafish model of TB infection, Roca *et al*. found that TNF induces reverse electron transport (RET) in mitochondrial complex I. This in turn drives the production of mitochondrial reactive oxygen species (mROS), causing macrophage necrosis. The complex I inhibitor metformin could be repurposed to inhibit TNF-induced mROS and necrosis in infected zebrafish and human macrophages, suggesting that this common antidiabetes drug may also be a useful adjunct therapy for TB. —STS

Science, abh2841, this issue p. 1431

CORONAVIRUS

Peaks of replacement

For the past 2 years, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has been tracked at fine resolution among health

care workers in the United Kingdom. Using the data from the REACT study of more than 100,000 participants, Elliott *et al*. show how the omicron variant BA.2 rapidly replaced its predecessor BA.1 in March 2022. Prevalence during March exceeded 6% and was the highest among older age groups and households with children. This led to the highest incidence seen in the pandemic thus far and to increased hospitalization rates. Meanwhile, a few recombinant BA.1/BA.2 infections were detected that need to be monitored. The UK government has recently halted funding for the REACT study, thus curtailing a source of unbiased long-term data on the trajectory of the SARS-CoV-2 pandemic and its health consequences. —CA

Science, abq4411, this issue p. 1432

RESEARCH ARTICLE SUMMARY

IMMUNOMETABOLISM

Tumor necrosis factor induces pathogenic mitochondrial ROS in tuberculosis through reverse electron transport

Francisco J. Roca, Laura J. Whitworth, Hiran A. Prag, Michael P. Murphy, Lalita Ramakrishnan*

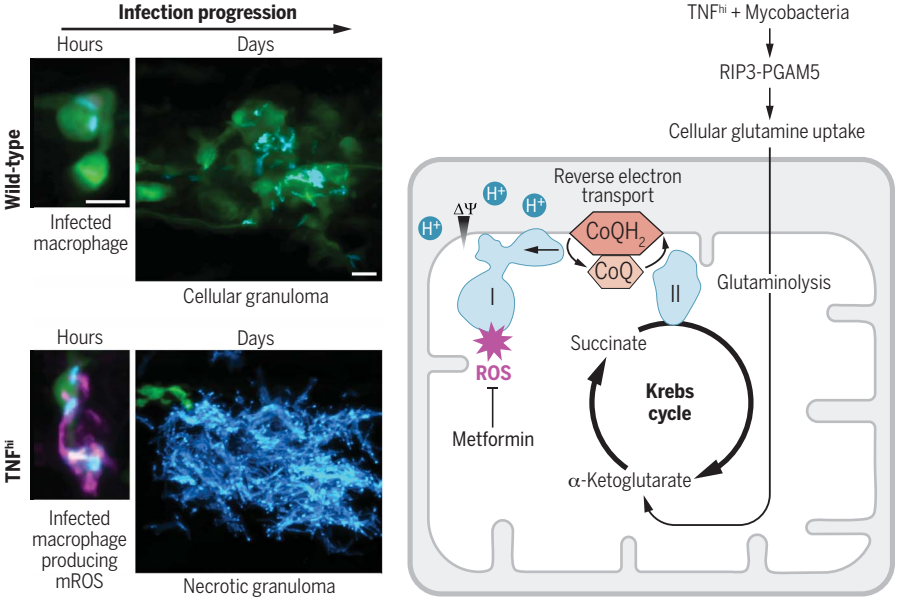
INTRODUCTION: Tumor necrosis factor (TNF) mediates resistance to tuberculosis. Excess TNF production, however, is detrimental because it induces pathogenic necrosis of infected macrophages in the tuberculous granuloma, which releases the mycobacteria into the extracellular milieu, promoting their growth and transmission to new hosts. Excess TNF, through the kinase RIP3 and the mitochondrial phosphatase PGAM5, increases reactive oxygen species such as superoxide and hydrogen peroxide in the mitochondria of mycobacterium-infected macrophages. These mitochondrial reactive oxygen species (mROS) initiate an elaborate interorganellar signaling circuit that ultimately causes macrophage necrosis and release of mycobacteria.

RATIONALE: How TNF signaling elevates mROS production is not known. To address this question in vivo, we used zebrafish larvae, taking advantage of their optical transparency and their amenability to genetic and pharmacological manipulation. Thus, we could visualize and quantify mROS and macrophage necrosis after these manipulations.

RESULTS: Typically, mROS are generated during normal respiration when electrons from reduced nicotinamide adenine dinucleotide (NADH), produced by metabolic pathways, enter the electron transport chain (ETC) and are transferred by forward electron transport from complex I to coenzyme Q (CoQ). We found that in wild-type animals without excess TNF, mycobacterial infection induced a small boost in multiple metabolic pathways that increased mROS through this process. This slight increase in mROS did not result in macrophage necrosis. In animals with excess TNF (TNF^{hi} animals), we found that the greatly increased mROS were not induced through conventional forward electron transport but rather through reverse electron transport (RET). RET occurs when increases in the pool of reduced CoQ ($CoQH_2$) from various metabolic pathways—in conjunction with a high proton-motive force across the mitochondrial inner membrane—cause electrons to flow back through complex I instead of forward into complex III. RET can generate large amounts of mROS at complex I. We found that increased oxidation of succinate at complex II was responsible for RET mROS and that this metabolite was the source of the accumulation of $CoQH_2$. Succinate is produced in the Krebs cycle, so we investigated its metabolic source. We found that TNF increased glutamine transport into the cell, boosting glutaminolysis, which increases the pool of α-ketoglutarate supplied to the Krebs cycle, resulting in increased succinate.

Mycobacteria play a critical role in TNF-induced necrosis at two distinct steps. They were required together with TNF to increase glutaminolysis and then again with the resultant mROS to induce necrosis. By contrast, TNF had no further role in the necrosis pathway beyond inducing mROS. Thus, virulent mycobacteria have evolved multiple orchestrated mechanisms to exploit host genetic vulnerabilities (i.e., dysregulated TNF levels) to mediate macrophage necrosis as a way of increasing the transmission that is critical to their survival. Delineation of the TNF-induced RET mROS pathway identified several drugs, already approved for other conditions, that inhibit the pathway at different steps. These drugs also inhibited TNF-induced macrophage necrosis and the animals' hypersusceptibility to infection.

CONCLUSION: RET, long thought to be an in vitro artifact, is now appreciated to play important homeostatic roles through moderate increases in mROS. However, excess RET has been shown to mediate the pathology associated with ischemia-reperfusion injury in heart attack and stroke. Our work shows that RET mROS also mediates tuberculosis pathology. Paradoxically, this means that the critical host determinant TNF can go from being protective to pathogenic depending on relative concentration, context, and the extent to which it can modulate host metabolism. Our prior work has shown that dysregulated TNF is also pathogenic in human tuberculosis. Therefore, the pathway-inhibiting drugs we have identified are promising host-targeting adjunctive drugs for tuberculosis, both drug-sensitive and drug-resistant. Metformin—a widely used, well-tolerated antidiabetic drug as well as a complex I inhibitor—is a particularly good candidate. ■



Excess TNF induces pathological mROS via RET in tuberculosis. Left: TNF induces mROS (magenta) in macrophages (green) infected with mycobacteria (blue), which causes their necrosis with exuberant growth of released mycobacteria in the debris of the tuberculosis granuloma. Right: Schematic depiction of how TNF works through RIP3 and PGAM5 to elevate mROS by RET through complex I. Metformin inhibits complex I to prevent TNF-induced mROS.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: lalitar@mrc-lmb.cam.ac.uk
Cite this article as F. J. Roca *et al.*, *Science* **376**, eabh2841 (2022). DOI: 10.1126/science.abh2841

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abh2841

RESEARCH ARTICLE

IMMUNOMETABOLISM

Tumor necrosis factor induces pathogenic mitochondrial ROS in tuberculosis through reverse electron transport

Francisco J. Roca[1]†, Laura J. Whitworth[1,2], Hiran A. Prag[3], Michael P. Murphy[1,3], Lalita Ramakrishnan[1,2]*

Tumor necrosis factor (TNF) is a critical host resistance factor against tuberculosis. However, excess TNF produces susceptibility by increasing mitochondrial reactive oxygen species (mROS), which initiate a signaling cascade to cause pathogenic necrosis of mycobacterium-infected macrophages. In zebrafish, we identified the mechanism of TNF-induced mROS in tuberculosis. Excess TNF in mycobacterium-infected macrophages elevates mROS production by reverse electron transport (RET) through complex I. TNF-activated cellular glutamine uptake leads to an increased concentration of succinate, a Krebs cycle intermediate. Oxidation of this elevated succinate by complex II drives RET, thereby generating the mROS superoxide at complex I. The complex I inhibitor metformin, a widely used antidiabetic drug, prevents TNF-induced mROS and necrosis of *Mycobacterium tuberculosis*–infected zebrafish and human macrophages; metformin may therefore be useful in tuberculosis therapy.

Tumor necrosis factor (TNF) is both a host resistance and susceptibility factor in tuberculosis (TB) (*1*–*3*). Findings in the genetically tractable and optically transparent zebrafish larva infected with *Mycobacterium marinum* (*Mm*) have revealed the mechanisms behind this dual effect (*4*–*6*). Although TNF is required for full microbicidal activity of mycobacterium-infected macrophages, its excess causes susceptibility by inducing their necrotic death, which releases mycobacteria into the growth-permissive extracellular environment (*4*, *7*–*9*). This pathogenic role of dysregulated TNF was revealed through a zebrafish forward genetic screen, which found that both a deficiency and an excess of leukotriene A4 hydrolase (LTA4H) cause susceptibility to *Mm* (*4*, *9*). LTA4H catalyzes the synthesis of the proinflammatory leukotriene B4 (LTB_4), and LTA4H/LTB_4 deficiency and excess produce TNF deficiency and excess, respectively (*4*). These zebrafish studies led to the identification of a common, functional human *LTA4H* variant associated with mortality from tuberculous meningitis, the severest form of TB (*4*, *10*). In cohorts in Vietnam and Indonesia, the high LTA4H–expressing variant was associated with increased cerebrospinal fluid TNF levels and increased mortality that was mitigated by adjunctive treatment with corticosteroids, broadly acting immunosuppressants (*3*, *4*, *10*). These findings implicated LTB_4 and TNF-induced inflammation in mortality (*3*, *4*, *10*). Moreover, high TNF concentrations were associated with mortality even among individuals without the high LTA4H–expressing variant; this suggested that TNF excess, resulting from diverse host genetic determinants, is a far-reaching host susceptibility factor in TB (*3*). Consistent with these findings, necrotic human tuberculous granulomas have more TNF than non-necrotic ones (*11*).

To gain mechanistic understanding of TNF-mediated pathogenic macrophage necrosis, we returned to the zebrafish larva. We found that excess TNF, acting through the kinase RIP3 and one of its substrates, PGAM5, increases mitochondrial reactive oxygen species (mROS) such as superoxide and hydrogen peroxide in infected macrophages (fig. S1) (*5*, *6*). These mROS activate an interorganellar signaling circuit that involves the lysosome and the endoplasmic reticulum. This ultimately causes mitochondrial calcium overload, which then leads to necrosis (fig. S1) (*5*, *6*). By exploiting the zebrafish larva's genetic and pharmacological tractability, we were able to determine how TNF induces pathogenic mROS in mycobacterium-infected macrophages.

TNF induces mROS through RET

Administering exogenous TNF to *Mm*-infected zebrafish larvae phenocopies genetically induced TNF excess, causing macrophage necrosis and susceptibility by 5 days after infection (fig. S2) (*4*). TNF selectively induces mROS in infected macrophages within 30 min, which rapidly triggers necrosis (*5*, *6*). Using a general mitochondria-targeted ROS and oxidative stress sensor, we found that in wild-type animals, *Mm* infection alone causes increases in mROS in infected macrophages by a factor of 1.7 to 2.2 relative to uninfected macrophages in the same animal (Fig. 1, A and B). In TNF^{hi} animals, mROS in infected macrophages were further increased by a factor of 3.6 to 6.6 relative to uninfected macrophages (Fig. 1, A and B). TNF did not increase mROS in uninfected macrophages, hence only infected macrophages were susceptible to TNF's effects (Fig. 1, A and B). Moreover, heat-killed *Mm* failed to induce mROS in both wild-type and TNF^{hi} animals, which suggested that an actively synthesized (or heat-labile) bacterial determinant is needed (Fig. 1B). To confirm that the increased TNF^{hi} mROS originated from superoxide production by the electron transport chain (ETC), we asked whether it was inhibited by compounds that disrupt mitochondrial electron transport. mROS were inhibited by four compounds that disrupt mitochondrial electron transport through distinct mechanisms (Fig. 1, C to F, fig. S2, and table S1). Thus, TNF-induced mROS originate in the ETC of mycobacterium-infected macrophages.

During normal respiration, complex I receives electrons from NADH (reduced nicotinamide adenine dinucleotide) and transfers them to CoQ (coenzyme Q), generating in the process a small amount of the mROS superoxide ($O_2^{\bullet-}$) through single electron donation to O_2 (*12*) (Fig. 2A). Increased $O_2^{\bullet-}$ production at complex I is generated by two distinct mechanisms (*12*). In the first, disruption of electron transfer, due to ETC damage or loss of cytochrome C during apoptosis, results in an accumulation of NADH derived from multiple metabolic pathways. When electrons from NADH enter complex I and cannot flow forward toward ubiquinone, they generate $O_2^{\bullet-}$ (Fig. 2A). In the second mechanism, increases in the extent of CoQ pool reduction ($CoQH_2$) from various metabolic pathways, in conjunction with a high proton-motive force across the mitochondrial inner membrane, cause electrons to flow back through complex I instead of forward into complex III (Fig. 2B) (*13*). This reverse electron transport (RET) by complex I generates mROS ($O_2^{\bullet-}$, which dismutates to H_2O_2) (Fig. 2B). These two mechanisms can be distinguished by the effects of the complex I inhibitor rotenone, which increases mROS from forward electron flow through complex I but reduces mROS from RET (Fig. 2, A and B, and table S1) (*12*, *14*). Rotenone increased mROS in the infected macrophages of wild-type animals, showing that they were generated by forward electron transport (Fig. 2C). By contrast, rotenone inhibited mROS in TNF^{hi} animals (Fig. 2D). Two other complex I inhibitors with different mechanisms of action also

[1]Molecular Immunity Unit, Cambridge Institute of Therapeutic Immunology and Infectious Diseases, Department of Medicine, University of Cambridge, Cambridge CB2 0AW, UK. [2]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. [3]MRC Mitochondrial Biology Unit, University of Cambridge, Cambridge CB2 0XY, UK.
*Corresponding author. Email: lalitar@mrc-lmb.cam.ac.uk
†Present address: Department of Biochemistry and Molecular Biology–B and Immunology, Biomedical Research Institute of Murcia (IMIB-Arrixaca), University of Murcia, 30120 Murcia, Spain.

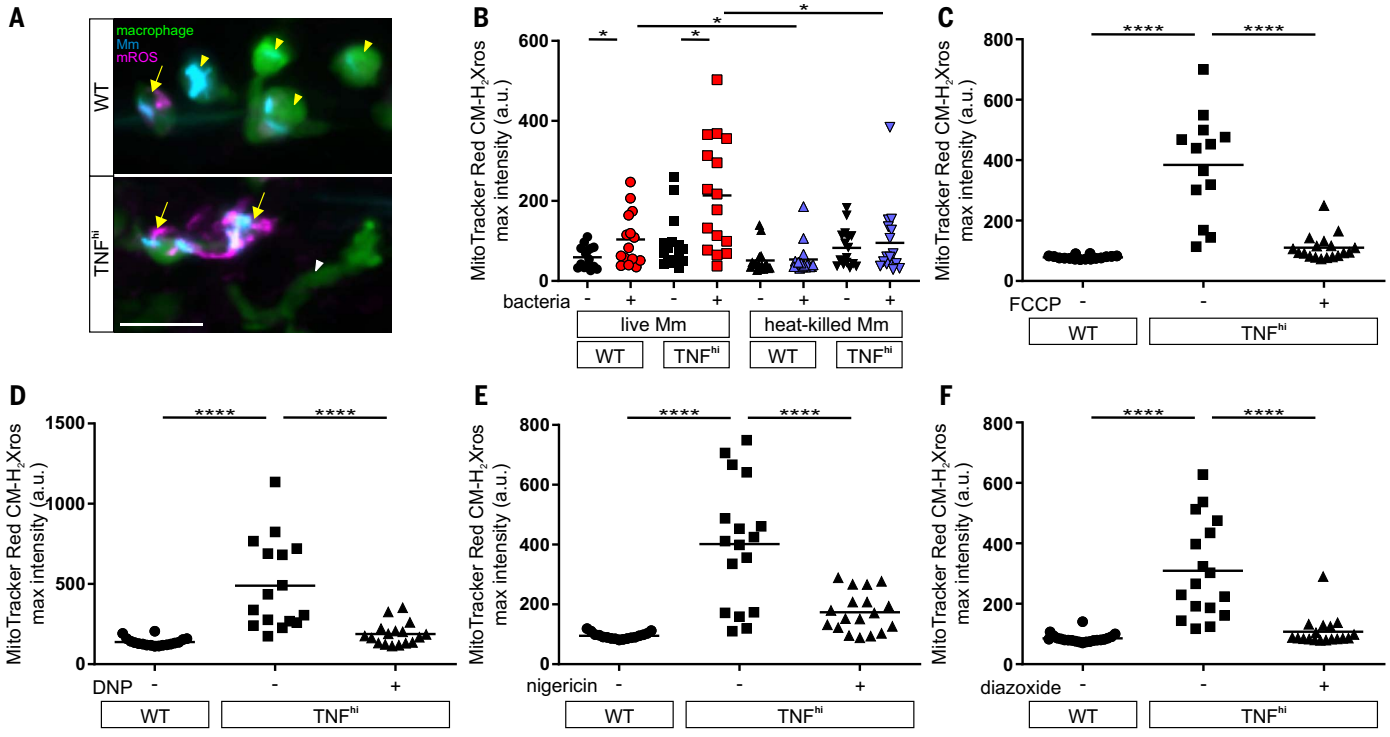


Fig. 1. ETC-derived mROS drive necrosis of *Mm*-infected macrophages in TNF-high conditions. (**A**) Representative pseudocolored confocal images of wild-type or TNF^{hi} larvae with yellow fluorescent protein (YFP)–expressing macrophages (green), 1 day after infection with EBFP2-expresssing *Mm* (blue), showing MitoTracker Red CM-H_2Xros (magenta) fluorescence. White arrowheads, uninfected macrophages; yellow arrowheads, infected macrophages; yellow arrows, infected macrophages positive for mROS. Scale bar, 20 μm. (**B**) Quantification of mROS in wild-type or TNF^{hi} larvae 9 hours after injection of live or heat-killed *Mm*. Each point represents the mean maximum intensity fluorescence of MitoTracker Red CM-H_2Xros per fish. Black symbols represent macrophages that do not contain bacteria. Red and purple symbols represent *Mm*-infected and heat-killed *Mm*-containing macrophages, respectively, in the same animal. Horizontal bars denote means; *P < 0.05 (one-way ANOVA with uncorrected Dunn's post-test for differences between macrophages in the same animal and with Tukey's post-test for differences between treatments). Data are representative of two independent experiments. (**C** to **F**) Quantification of mROS in larvae 1 day after infection with *Mm* that are wild-type or TNF^{hi} treated with FCCP (C), DNP (D), nigericin (E), or diazoxide (F), or vehicle. Horizontal bars denote means; ****P < 0.0001 (one-way ANOVA with Tukey's post-test). Data are representative of two or three independent experiments.

inhibited TNF^{hi} mROS (fig. S3 and table S1). Thus, TNF^{hi} mROS are generated by RET rather than by forward electron transport.

To corroborate that RET was responsible for TNF^{hi} mROS, we expressed *Ciona intestinalis* alternative oxidase (AOX) in TNF^{hi} larvae through injection of its mRNA. AOXs, which are present in plants, fungi, and some invertebrates but are absent in vertebrates, catalyze the transfer of electrons from the $CoQH_2$ pool directly to O_2, bypassing complexes III and IV (fig. S4A) (*15*). AOX has been shown to prevent excessive reduction of the CoQ pool and mROS increases from RET (fig. S4A) (*15*). Thus, if TNF-induced mROS are generated by RET, they should be prevented by AOX expression (*15*). We confirmed that the *C. intestinalis* AOX was active in zebrafish by showing that AOX-expressing animals were resistant to cyanide, which poisons the ETC by inhibiting complex IV (fig. S4B) (*15*, *16*). AOX expression decreased TNF^{hi} mROS, consistent with generation by RET from a reduced CoQ pool (Fig. 2, B and E).

Finally, we extended these findings to *M. tuberculosis* (*Mtb*), the agent of human TB, using a leucine and pantothenic acid *Mtb* auxotroph that can be used in containment level 2 facilities (*6*). *Mtb* produced increases in mROS similar to those produced by *Mm* (Fig. 2F). Moreover, rotenone increased mROS in wild-type macrophages and inhibited TNF-induced mROS (Fig. 2, G and H). Thus, TNF-induced mROS increases in *Mtb*-infected macrophages are also derived from RET.

Although there are multiple sources of increased $CoQH_2$, the most compelling candidate from both in vitro and in vivo studies was the increased oxidation of succinate at complex II (Fig. 2B) (*13*, *17*). We tested this idea using three complex II inhibitors: atpenin A, thenoyltrifluoroacetone (TTFA), and dimethyl malonate (DM-malonate), which is a prodrug of the competitive succinate dehydrogenase inhibitor malonate. All three inhibitors abolished mROS (Fig. 3A and table S1). If increased succinate oxidation at complex II were the source of RET and mROS, then increasing the mitochondrial succinate pool should have induced mROS even in wild-type animals in the absence of TNF^{hi} conditions. Diethyl succinate—a cell-permeable succinate ester known to increase mitochondrial succinate concentrations (*18*)—increased mROS in macrophages of wild-type animals (Fig. 3B). Diethyl butylmalonate (DEBM)—an inhibitor of the mitochondrial succinate transporter, which causes accumulation of endogenous mitochondrial succinate (*19*)—performed similarly (Fig. 3B). Thus, increased oxidation of succinate at complex II is necessary and sufficient for TNF-induced RET and mROS.

TNF-activated glutaminolysis increases mitochondrial succinate

We next investigated the metabolic source of the increased succinate concentration. Increased glycolysis, fatty acid oxidation, and glutaminolysis can all increase succinate by increasing Krebs cycle activity through increased input of pyruvate, acetyl-CoA, and α-ketoglutarate, respectively (Fig. 4A). We

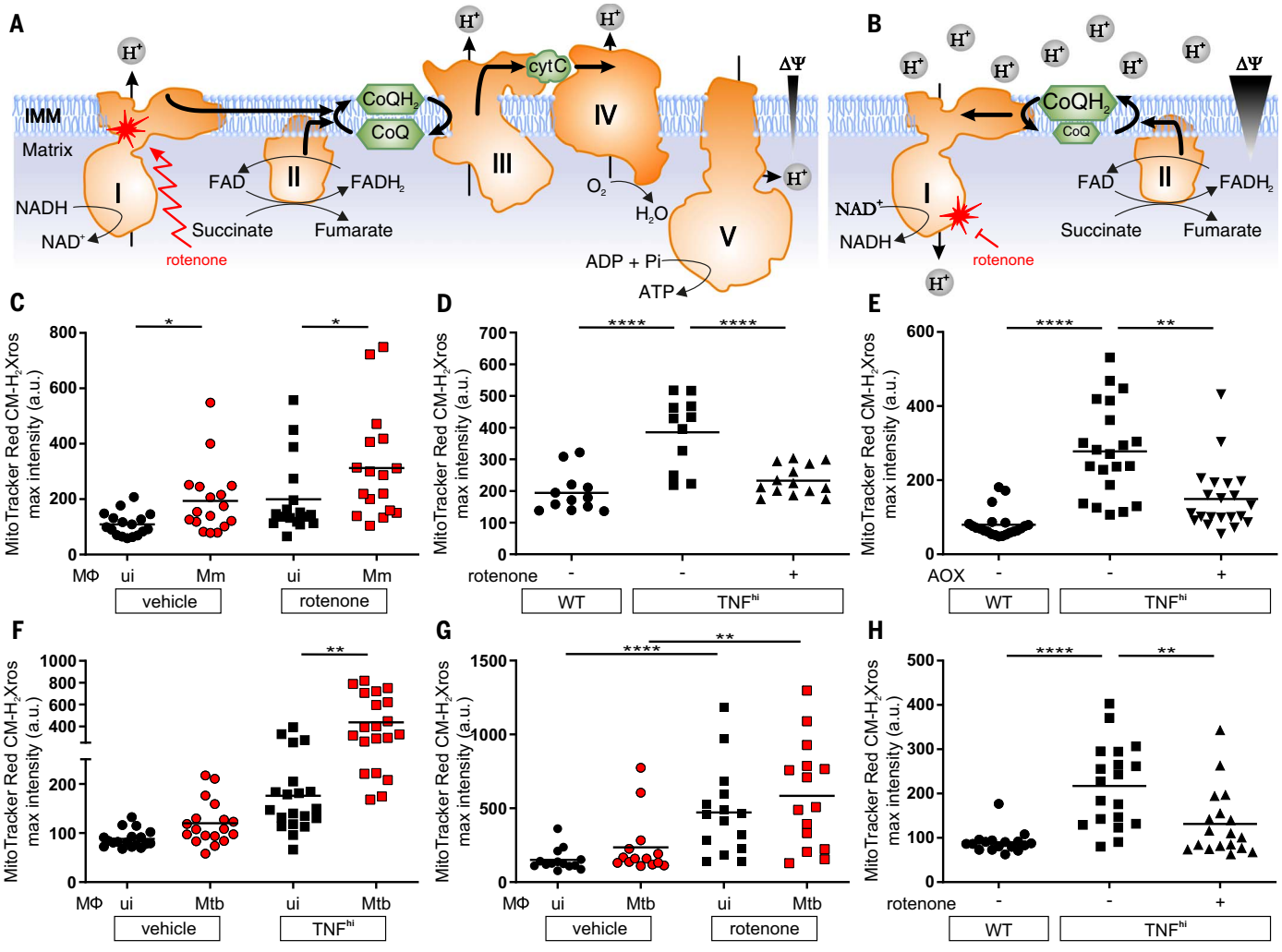


Fig. 2. TNF induces RET mROS at complex I in mycobacterium-infected macrophages. (**A** and **B**) Illustrations of mROS production at complex I during forward electron transport (A) and reverse electron transport (B). FAD, flavin adenine dinucleotide; ΔΨ, membrane potential; IMM, inner mitochondrial membrane; ADP/ATP, adenosine diphosphate/triphosphate; cytC, cytochrome C; I to V, complexes I to V; zigzag arrows, induction; red blunted arrows, inhibition. (**C** to **H**) Quantification of mROS in larvae 1 day after infection with *Mm* [(C) to (E)] or *Mtb* [(F) to (H)] that are wild-type treated with rotenone or vehicle (C); wild-type or TNF^{hi} treated with rotenone or vehicle (D); wild-type, TNF^{hi}, or TNF^{hi} expressing AOX (E); wild-type or TNF^{hi} (F); wild-type treated with rotenone or vehicle (G); or wild-type or TNF^{hi} treated with rotenone or vehicle (H). Horizontal bars denote means; *P < 0.05, **P < 0.01, ****P < 0.0001 [one-way ANOVA with Dunn's post-test in (C), (G), and (H); Tukey's post-test in (D) and (E); uncorrected Dunn's post-test in (F)]. Black and red symbols in (C), (F), and (G) represent uninfected (ui) and infected macrophages, respectively, in the same animals. Data are representative of two or three independent experiments [(C to G)] or a single experiment (H).

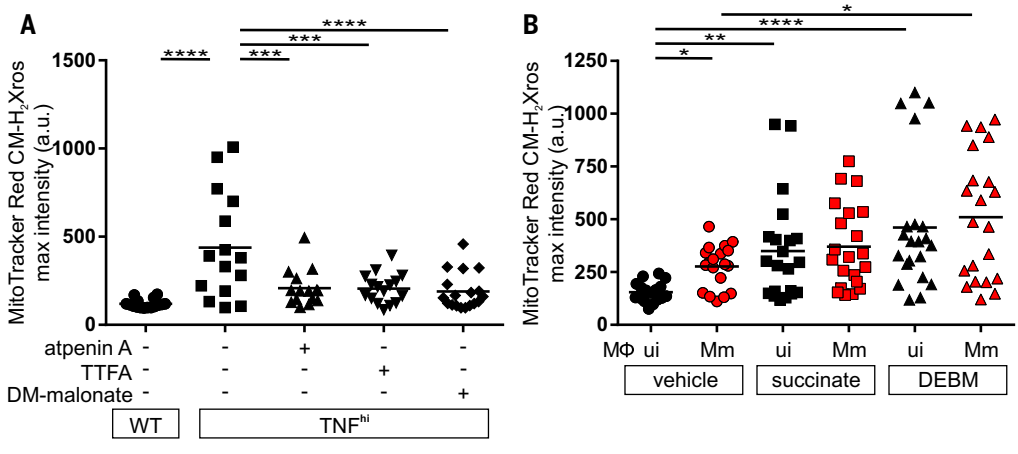


Fig. 3. TNF increases succinate-dependent mROS in mycobacterium-infected macrophages. (**A** and **B**) Quantification of mROS in larvae 1 day after infection with *Mm* that are wild-type or TNF^{hi} treated with atpenin A, TTFA, DM-malonate, or vehicle (A) or wild-type treated with succinate, DEBM, or vehicle. Horizontal bars denote means; *P < 0.05, **P < 0.01, ***P < 0.001,****P < 0.0001 [one-way ANOVA with Tukey's post-test (A) or Dunn's post-test (B)]. Black and red symbols in (B) represent uninfected (ui) and *Mm*-infected (*Mm*) macrophages, respectively, in the same animal. Data are representative of two or three independent experiments.

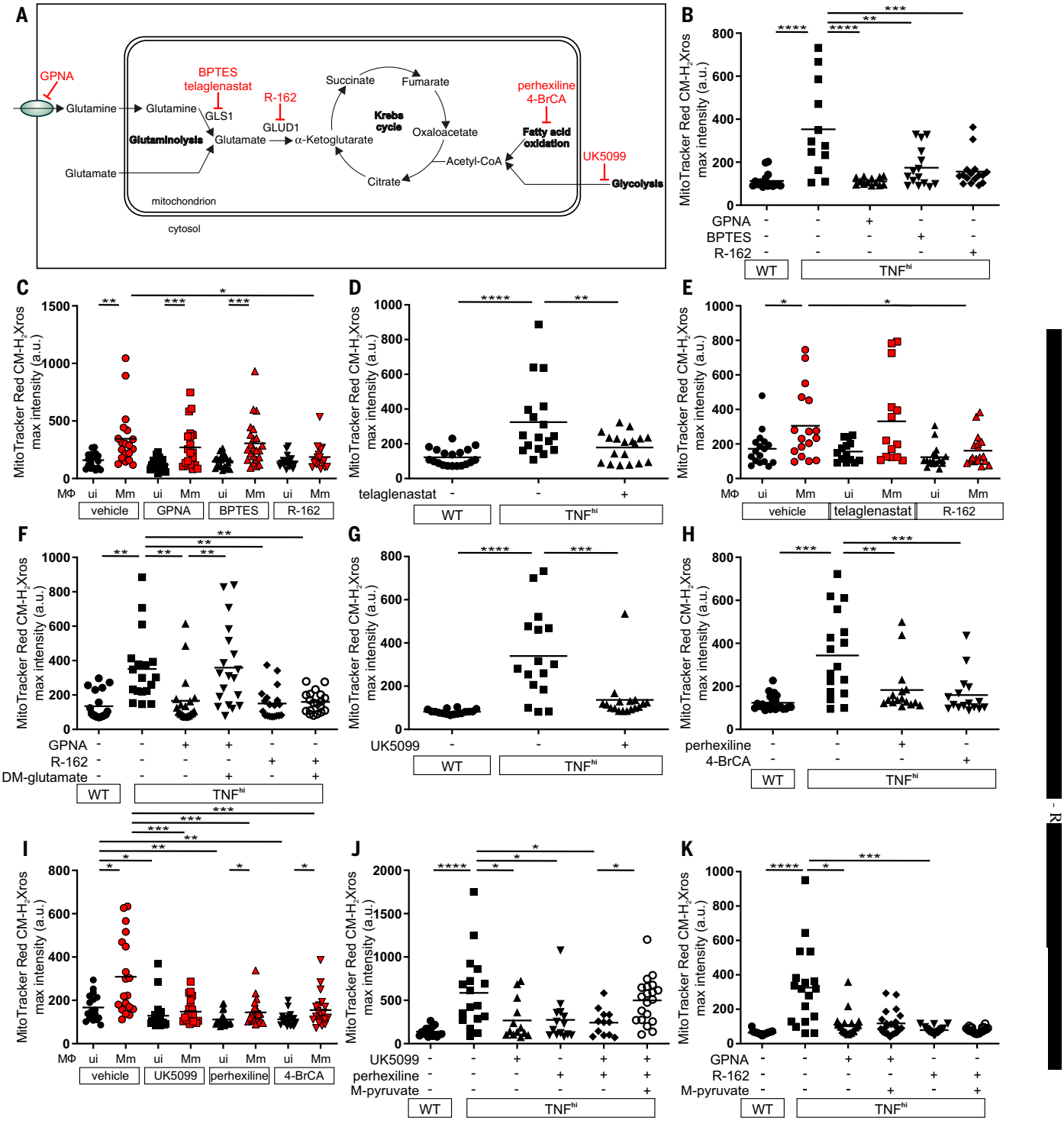


Fig. 4. TNF-induced glutamine cellular uptake and increased glutaminolysis is responsible for RET and mROS production in mycobacterium-infected macrophages. (**A**) Illustration of main metabolic pathways fueling the Krebs cycle with inhibitors used (truncated red arrows). GLUD1, glutamate dehydrogenase 1. (**B** to **K**) Quantification of mROS in larvae 1 day after infection with *Mm* that are wild-type or TNF^{hi} treated with GPNA, BPTES, R-162, or vehicle (B); wild-type treated with GPNA, BPTES, R-162, or vehicle (C); wild-type or TNF^{hi} treated with telaglenastat or vehicle (D); wild-type treated with telaglenastat, R-162, or vehicle (E); wild-type or TNF^{hi} treated with vehicle, or GPNA or R-162 alone or in combination with DM-glutamate (F); wild-type or TNF^{hi} treated with UK5099 or vehicle (G); wild-type or TNF^{hi} treated with perhexiline, 4-BrCA, or vehicle (H); wild-type treated with UK5099, perhexiline, 4-BrCA, or vehicle (I); wild-type or TNF^{hi} treated with vehicle, or UK5099 or perhexiline alone or in combination with M-pyruvate (J); or GPNA or R-162 alone or in combination with M-pyruvate (K). Horizontal bars denote means; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$ [one-way ANOVA with Tukey's post-test in (B), (D), (F) to (H), (J), and (K); Dunn's post-test in (C), (E), and (I)]. Black and red symbols in (C), (E), and (I) represent uninfected (ui) and *Mm*-infected (*Mm*) macrophages, respectively, in the same animals. Data are representative of two or three independent experiments [(B) to (D), (G) to (I)] or a single experiment [(E), (F), (J), and (K)].

focused on glutaminolysis, which has been linked to TNF-mediated cell death (*20*, *21*). Glutamine, the major amino acid transported in the circulation, is taken up into cells by multiple glutamine transporters and then into mitochondria, where it is converted to glutamate and then α-ketoglutarate in the Krebs cycle (Fig. 4A). Four potential plasma membrane transporters that contribute to cellular glutamine uptake are highly expressed in human and zebrafish monocytes and macrophages (*22*–*24*). Of these, SLCA15 and SLC38A2 were identified in a screen for proteins phosphorylated by the RIP3 kinase in the context of necroptosis, a different form of TNF-mediated programmed cell death (*25*). Although distinct from necroptosis, our macrophage necrosis pathway also features RIP3, which is required for TNF-mediated mROS induction in mycobacterium-infected macrophages (fig. S1) (*5*, *6*). Therefore, we tested L-γ-glutamyl-p-nitroanilide (GPNA), an inhibitor of both transporters (table S1). GPNA inhibited mROS in TNF^{hi} macrophages without affecting mROS in wild-type macrophages (Fig. 4, A to C). We therefore hypothesized that TNF-RIP3–activated glutamine transport is the specific source of the increased mitochondrial glutamine for increased glutaminolysis, thereby increasing succinate. If correct, blocking the conversion of glutamine to glutamate should also specifically block TNF^{hi} mROS. Bis-2-(5-phenylacetamido-1,3,4-thiadiazol-2-yl)ethyl sulfide (BPTES) and CB-839 (telaglenastat), two different inhibitors of glutaminase 1 (GLS1), performed as expected, inhibiting mROS in TNF^{hi} but not wild-type macrophages (Fig. 4, A to E, and table S1). By contrast, when the conversion of glutamate to α-ketoglutarate was inhibited using R-162, mROS was inhibited in both TNF^{hi} and wild-type macrophages (Fig. 4, A, B, C, and E, and table S1). Thus, although the smaller increase in mROS from infection alone also requires glutaminolysis, it can be sustained by mitochondrial glutamate transported directly from the cytosol where it is produced through transamination reactions (Fig. 4A, C and E). Finally, to confirm the specificity of GPNA and R-162 in our system, we used each inhibitor in combination with dimethyl glutamate (DM-glutamate), a cell-permeable source of glutamate (table S1). Dimethyl glutamate restored GPNA-inhibited mROS but not R-162–inhibited mROS (Fig. 4F). Thus, TNF stimulation of infected macrophages specifically activates glutamine uptake to increase glutaminolysis to induce mROS.

If high TNF also increases glycolysis and/or fatty acid oxidation (Fig. 4A), then inhibiting these pathways should also specifically inhibit TNF^{hi} but not wild-type mROS. However, inhibition of mitochondrial pyruvate transport using UK5099 removed mROS in both wild-type and TNF^{hi} animals, as did inhibition of fatty acid oxidation using perhexiline or 4-bromocrotonic acid (4-BrCA) (Fig. 4, G to I, and table S1). We confirmed the specificity of UK5099 and perhexiline by showing that methyl pyruvate (M-pyruvate), a cell-permeable pyruvate derivative, restored mROS inhibited by them but not by GPNA or R-162 (Fig. 4, J and K, and table S1). Thus, TNF and infection together activate cellular glutamine uptake, and the resultant increase in glutaminolysis is the specific source of the increased succinate. Because oxidation of excess succinate would increase the concentrations of the downstream intermediates malate and oxaloacetate (a potent complex II inhibitor) (*26*), glycolysis and fatty acid oxidation would be required to play a "supporting role" by providing acetyl-CoA to consume oxaloacetate. Thus, the build-up of oxaloacetate would be prevented, allowing continued complex II activity (Fig. 4A).

We used liquid chromatography–mass spectrometry to quantify succinate concentrations in the larvae under the different conditions. Infection and TNF combined (but neither alone) increased succinate concentrations over baseline (Fig. 5A and data S1). Moreover, GPNA and BPTES inhibited this increase, as predicted (Fig. 5A and data S1). Although further validation of the source of succinate by measurement of flux to it from stable isotope–labeled precursors such as glutamine was not technically possible in this in vivo system, our findings that both mROS and succinate concentrations increase in the TNF^{hi} state and decrease to wild-type levels upon inhibiting glutamine uptake or its conversion to glutamate provide strong evidence that glutaminolysis from increased glutamine transport is the source of the increased succinate. As with mROS increases, these succinate increases also occurred rapidly within 30 min of TNF administration. The rapid induction of succinate and mROS is consistent with TNF-RIP3–induced posttranslational modifications (e.g., phosphorylation), as previously proposed (*20*, *21*). Accordingly, RIP3 knockdown inhibited TNF-induced succinate in infected animals (Fig. 5B and data S1). Finally, TNF-induced succinate was also inhibited by knockdown of PGAM5, a mitochondrial phosphatase, which is required together with RIP3 both for TNF-mediated necroptosis (*27*) and for TNF-induced mROS and necrosis of mycobacterium-infected macrophages in our pathway (Fig. 5B, data S1, and fig. S1) (*5*, *6*). Thus, TNF signals via RIP3 and PGAM5 to activate glutamine transport to increase glutaminolysis and Krebs cycle succinate.

TNF, mROS, and mycobacteria play discrete roles in macrophage necrosis

We dissected the interactions among TNF, mROS, and mycobacteria and their roles at distinct steps of the pathway. We had shown that both TNF and mycobacteria are required to increase mitochondrial succinate, which is required to induce mROS. Because exogenous succinate could induce mROS in wild-type animals in both infected and uninfected macrophages (Fig. 3B), we concluded that the only role for TNF and mycobacteria in mROS induction in this system is to increase mitochondrial succinate.

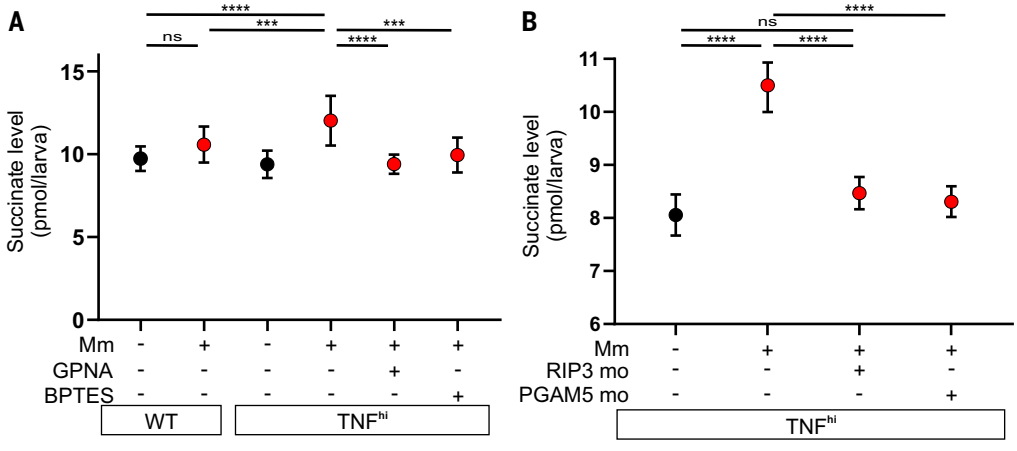


Fig. 5. TNF-induced glutaminolysis increases succinate levels in mycobacterium-infected macrophages in a RIP3- and PGAM5-dependent manner. (**A** and **B**) Quantification of succinate in zebrafish larvae 1 day after infection with *Mm* or mock-injected, that are wild-type or TNF^{hi} treated with GPNA, BPTES, or vehicle (A) and TNF^{hi}, TNF^{hi} RIP3 morphants, or TNF^{hi} PGAM5 morphants (B). Each point represents the mean of four independent experiments in (A) and two independent experiments in (B). Vertical bars denote pooled SD. ***P < 0.001, ****P < 0.0001 (one-way ANOVA with Tukey's post-test); ns, not significant.

We previously showed that mROS are required for macrophage necrosis (*5*). We now asked whether they were sufficient to complete macrophage necrosis, or whether TNF and/or mycobacteria were further required downstream of mROS induction. Macrophage necrosis results in exuberant extracellular mycobacterial growth in characteristic cords (Fig. 6A). Bacterial cording can be used as a reliable surrogate marker for infected macrophage death (*9*). We found that both exogenous succinate and DEBM induced the necrosis of infected macrophages, as evidenced by increased bacterial cording (Fig. 6B). This necrosis was a direct consequence of RET mROS

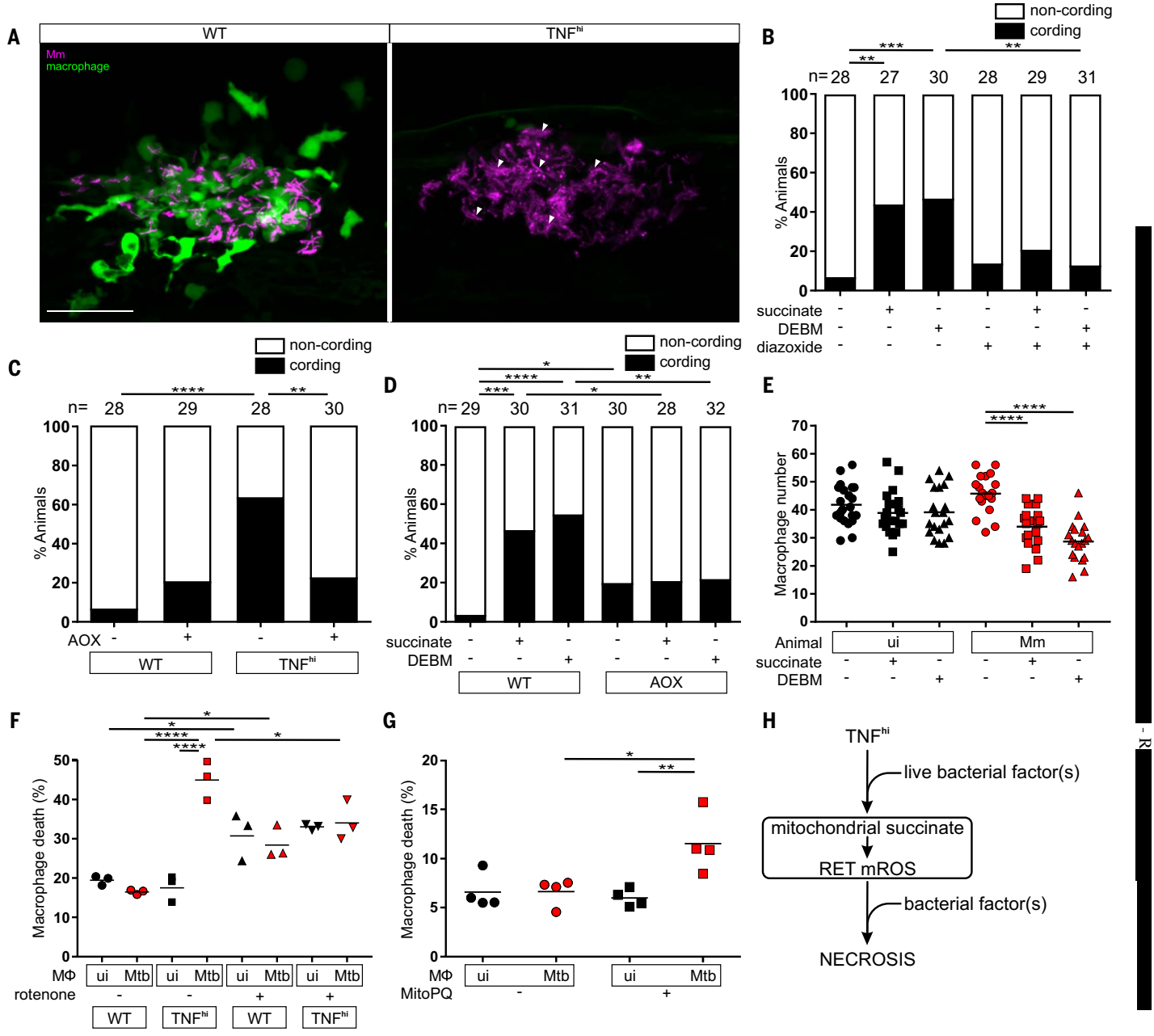


Fig. 6. TNF-mediated increased glutamine cellular uptake in mycobacterium-infected macrophages increases succinate oxidation, mROS, and necrosis. (**A**) Representative pseudocolored confocal images of 5-dpi granulomas in wild-type or TNF^{hi} larvae with YFP-expressing macrophages (green) infected with tdTomato-expressing *Mm* (magenta). Arrowheads, extracellular cording bacteria. Scale bar, 50 μm. (**B**) Bacterial cording in wild-type larvae 5 days after infection with *Mm*, treated with vehicle, or succinate or DEBM alone or in combination with diazoxide; $^{**}P < 0.01$, $^{***}P < 0.001$ (Fisher's exact test). (**C**) Bacterial cording 5 days after infection with *Mm* in wild-type and TNF^{hi} larvae and wild-type and TNF^{hi} larvae expressing AOX; $^{**}P < 0.01$, $^{****}P < 0.0001$ (Fisher's exact test). (**D**) Bacterial cording 5 days after infection with wild-type or AOX-expressing larvae infected with *Mm* and treated with succinate, DEBM, or vehicle; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$ (Fisher's exact test). (**E**) Number of trunk macrophages in *Mm*-infected larvae and mock-injected (ui) larvae 1 day after infection. Horizontal bars denote means; $^{****}P < 0.0001$ (one-way ANOVA with Dunn's post-test). (**F** and **G**) Percentage of dead THP-1 macrophages 5 hours after addition of TNF, treated with rotenone or vehicle starting 1 hour before TNF addition (F) or MitoParaquat (MitoPQ) or vehicle for 5 hours (G). Black and red symbols represent uninfected (ui) and *Mtb*-infected macrophages, respectively, within the same treatment well. Horizontal bars denote means; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{****}P < 0.0001$ (one-way ANOVA with Tukey's post-test). (**H**) Schematic diagram showing the role of TNF, mROS, and mycobacterial factor(s) in TNF-mediated necrosis of mycobacterium-infected macrophages. Data are representative of two independent experiments [(C), (D), (E), and (G)] or a single experiment [(B) and (F)].

production, as disrupting the ETC with diazoxide reduced cording (Fig. 6B). Moreover, bypassing the ETC by AOX expression—which decreased TNF-induced mROS (Fig. 2E)—inhibited TNF-mediated macrophage necrosis (Fig. 6C) as well as succinate- and DEBM-induced necrosis in wild-type animals (Fig. 6D). Thus, TNF plays no further role in macrophage necrosis beyond increasing mitochondrial succinate.

To determine whether mycobacteria were required for necrosis downstream of mROS induction, we examined whether diethyl succinate and DEBM could also kill uninfected macrophages by enumerating macrophages in infected and uninfected animals (*5*). Diethyl succinate and DEBM reduced macrophage numbers only in the infected animals; this finding suggests that unlike TNF, mycobacteria are required downstream of mROS to induce necrosis (Fig. 6E and fig. S2).

Similar results were observed in human macrophages derived from the monocytic cell line THP-1. We previously showed that TNF induces necrosis in *Mtb*-infected THP-1 cells through the same interorganellar pathway downstream of mROS as in *Mm*-infected zebrafish (fig. S2) (*6*). We used rotenone to confirm that RET was responsible for mROS induction in these cells. In the absence of TNF, rotenone increased death of both infected and uninfected cells, as expected from the oxidative stress it induces, but we found a specific reduction of TNF-induced death of infected macrophages (Fig. 6F). Next, to test our findings from zebrafish concerning the role of TNF, mROS, and mycobacteria, we treated *Mtb*-infected THP-1 cells with MitoParaquat (MitoPQ), a mitochondria-targeted compound that produces superoxide through redox cycling at the complex I flavin site (table S1). MitoPQ increased necrosis in the absence of TNF but only in infected macrophages (Fig. 6G and fig. S2). This confirmed that TNF has no further role in the necrosis pathway beyond inducing mROS, whereas mycobacteria are required downstream of mROS induction. One or more mycobacterial factors shared between *Mm* and *Mtb* operate at two distinct points in this pathway: first to enable TNF-mediated mROS by activating cellular glutamine uptake and increasing mitochondrial succinate to produce complex II–mediated RET-ROS, and then to promote the necrosis of macrophages experiencing this mROS (Fig. 6H).

The mROS pathway reveals host-targeting drugs for TB

We previously showed that blocking mROS using scavengers such as *N*-acetylcysteine inhibited TNF-induced macrophage necrosis and restored resistance (*5*). Four of the compounds used here to inhibit mROS, and thus to delineate the mechanism of mROS production, are approved oral drugs or are under investigation for other conditions. We therefore assessed whether these drugs also inhibited macrophage necrosis (fig. S5). These included diazoxide, a disruptor of electron transport that is approved for hyperinsulinemic hypoglycemia; perhexiline, a mitochondrial carnitine palmitoyltransferase-1 inhibitor that is approved for angina; telaglenastat, a GLS1 inhibitor that is in clinical trials for cancer; and DM-malonate, the complex II inhibitor that has been shown to prevent ischemia-reperfusion injury in models of heart attack (fig. S5 and table S1) (*28*). All four inhibited TNF-mediated macrophage necrosis in zebrafish (Fig. 7, A to E). We then asked whether metformin, a widely used antidiabetic drug that inhibits complex I (fig. S5 and table S1) (*29*), could be a potential host-targeting drug to prevent TNF-induced pathogenic macrophage necrosis in TB. Metformin inhibited TNF-elicited mROS in *Mm*-infected larvae, as did its more hydrophobic derivative, phenformin (Fig. 7F). Metformin also inhibited TNF-mediated necrosis of *Mm*-infected macrophages (Fig. 7G). Moreover, it also inhibited *Mm*-infected macrophage necrosis resulting from increased mitochondrial succinate (Fig. 7H). Thus, although metformin has pleiotropic effects and is a relatively weak complex I inhibitor (*29*), it specifically inhibits TNF-mediated necrosis by blocking RET-generated mROS at complex I. Finally, metformin inhibited mROS in the infected macrophages of *Mtb*-infected zebrafish (Fig. 7I) and inhibited necrosis of *Mtb*-infected THP-1 cells (Fig. 7J), thereby confirming that its inhibitory activity was relevant in the context of *Mtb* infection.

Discussion

Although long thought to be an in vitro artifact, moderate levels of RET and resultant increases in mROS have important homeostatic roles in cell differentiation and oxygen sensing (*13*). However, excess RET has pathological roles in ischemia-reperfusion injury of the heart and brain (*18*, *30*). During ischemia, rewiring of the Krebs cycle reduces fumarate concentrations, leading to succinate accumulation (*18*). During the reperfusion phase, rapid oxidation of the accumulated succinate triggers RET and mROS, which causes tissue necrosis leading to irreparable organ damage (*18*, *30*). The TNF-mediated necrosis pathway described here has two important differences. First, the source of the succinate is different; second, in ischemia-reperfusion injury, the mROS alone appear sufficient to drive necrosis, whereas a second "hit" in the form of one or more bacterial determinants is required in our TNF-induced macrophage necrosis pathway. Perhaps the inflammatory milieu generated during ischemia generates the additional signal(s) that combine with mROS to cause necrosis.

Work using cultured macrophages has shown that succinate is responsible for generating proinflammatory responses to lipopolysaccharide (LPS), a key virulence determinant of Gram-negative bacteria (*19*). LPS causes macrophages to switch to aerobic glycolysis while generating succinate from enhanced glutaminolysis by an undescribed means. Succinate induces mROS, likely through RET, and these mROS drive proinflammatory cytokines via hypoxia-inducible factor–1α stabilization (*19*, *31*). This sequence contrasts with the pathway described here, where TNF is upstream, not downstream, of mROS and TNF is not among the cytokines induced by LPS and succinate. Thus, distinct pathogenic determinants specific to Gram-negative bacteria and mycobacteria—a cell wall constituent versus a product of live mycobacteria—channel mROS to produce discrete cellular responses.

We were particularly interested in pursuing this TNF-mediated necrosis pathway because of its clinical implications. Currently, tuberculous meningitis is treated with adjunctive corticosteroids, which are broadly immunosuppressive and have multiple additional serious adverse effects. Our prior studies on the TNF-mediated necrosis pathway identified several pathway-specific drugs that inhibit macrophage necrosis without being broadly anti-inflammatory, all with a decades-long history of use in humans for other conditions (*5*, *6*). This work now identifies additional drugs, including the widely used oral antidiabetic drug metformin. Metformin readily crosses the blood-brain barrier, resulting in high brain and cerebrospinal fluid concentrations (table S1) (*32*). This highlights its potential therapeutic utility in tuberculous meningitis. Metformin was reported to ameliorate *Mtb* infection in mice via diverse mechanisms, including broadly acting anti-inflammatory effects, and to enhance the efficacy of antitubercular antibiotics in one but not another study, leading to an ongoing trial as an adjunctive agent for lung TB (*33*–*36*). Adjunctive corticosteroid treatment has been suggested to reduce inflammation and bacterial burdens in lung TB, the most common, contagious form that sustains the global disease burden (*37*, *38*). It will be interesting to see whether metformin particularly benefits individuals with the high-*LTA4H* genotype. Moreover, given the association of TNF with necrotic lung granulomas (*11*), metformin may increase resolution of necrotic lesions.

Materials and methods

Zebrafish husbandry and infections

Zebrafish husbandry and experiments were conducted in compliance with guidelines from the UK Home Office using protocols approved

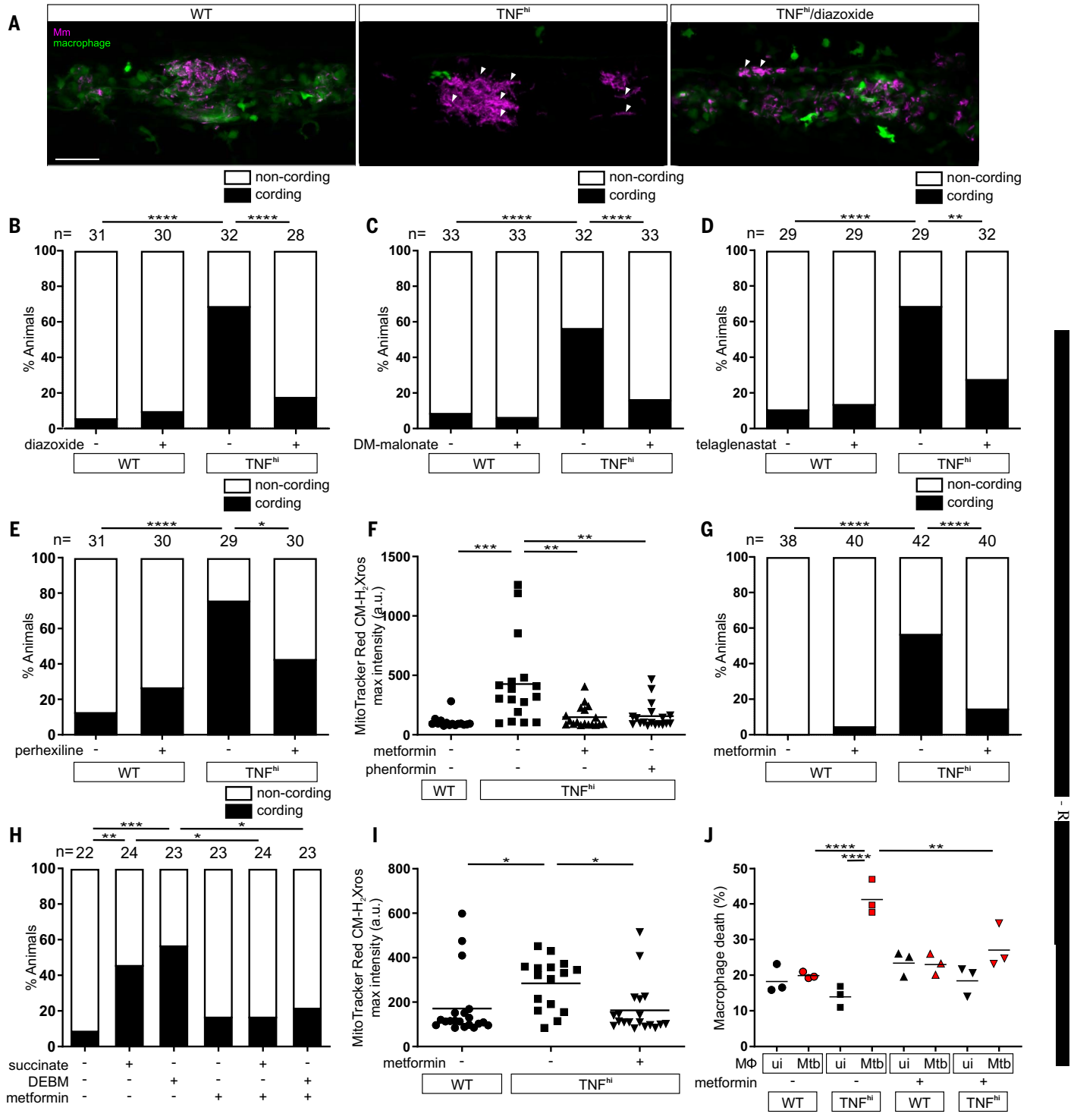


Fig. 7. Currently available drugs can intercept TNF-induced mROS production and inhibit necrosis of mycobacterium-infected macrophages. (**A**) Representative pseudocolored confocal images of 5-dpi granulomas in larvae with YFP-stained macrophages (green) that are wild-type or TNF^{hi} treated with diazoxide or vehicle, infected with red fluorescent *Mm* (magenta). Arrowheads, extracellular cording bacteria. Scale bar, 50 μm. (**B** to **E**) Bacterial cording in wild-type or TNF^{hi} larvae 5 days after infection with *Mm*, treated with vehicle or diazoxide (B), DM-malonate (C), telaglenastat (D), or perhexiline (E). *P < 0.05, **P < 0.01, ****P < 0.0001 (Fisher's exact test). (**F**) Quantification of mROS in wild-type or TNF^{hi} larvae 1 day after infection with *Mm*, treated with metformin, phenformin, or vehicle. Horizontal bars denote means; **P < 0.01, ***P < 0.001 (one-way ANOVA with Tukey's post-test). (**G**) Bacterial cording in wild-type or TNF^{hi} larvae 5 days after infection with *Mm*, treated with metformin or vehicle. ****P < 0.0001 (Fisher's exact test). (**H**) Bacterial cording in wild-type larvae 5 days after infection with *Mm*, treated with vehicle or with succinate or DEBM alone or in combination with metformin. *P < 0.05, **P < 0.01, ***P < 0.001 (Fisher's exact test). (**I**) Quantification of mROS in wild-type or TNF^{hi} 1 day after infection with *Mtb*, treated with metformin or vehicle. Horizontal bars denote means; *P < 0.05 (one-way ANOVA with Tukey's post-test). (**J**) Percentage

of dead THP-1 macrophages at 5 hours after addition of TNF, treated with metformin or vehicle starting 1 hour before TNF addition. Black and red symbols represent uninfected (ui) and *Mtb*-infected macrophages, respectively, within the same treatment well. Horizontal bars denote means; $^{**}P < 0.01$, $^{****}P < 0.0001$ (one-way ANOVA with Tukey's post-test). Data are representative of two independent experiments [(B) to (G) and (I)] or a single experiment [(H) and (J)].

by the Animal Welfare and Ethical Review Body of the University of Cambridge. Zebrafish AB wild-type strain (Zebrafish International Resource Center) (ZFIN ID: ZDB-GENO-960809-7) and the transgenic line *Tg(mpeg1:YFP)*^{w200} (with yellow fluorescent macrophages) (ZFIN ID: ZDB-FISH-150901-6828) (*6*) in the AB background were used. All zebrafish lines were maintained in buffered reverse osmotic water systems as described (*6*). Zebrafish embryos were housed at 28.5°C in fish water from collection to 1 day post-fertilization (dpf) and in E2 Embryo Medium diluted to 0.5× (E2/2) supplemented with 0.003% 1-phenyl-2-thiourea (PTU) (Sigma) from 1 dpf to prevent pigmentation (*6*). Larvae (of undetermined sex given the early developmental stages used) were anesthetized, infected at 2 dpf via caudal vein (CV) injection for all assays, and randomly allotted to the different experimental conditions as described (*6*, *39*). Sample size was determined based on previous similar experiments or on pilot experiments.

Bacterial strains

Mm M strain (ATCC #BAA-535) and *Mtb* H37Rv strain, mc^{2}6206 Δ*leuD* Δ*panCD* (*40*) expressing tdTomato, mWasabi, or EBFP2 were grown as described (*39*, *41*). For experiments to assay bacterial cording and number of macrophages in the trunk of the animal, zebrafish larvae were infected with 150 to 200 tdTomato-expressing *Mm*. To assess mROS, larvae were infected with 90 to 120 EBFP2-expressing or 84 mWasabi-expressing *Mm*, 80 to 100 EBFP2-expressing *Mtb*, or injected with 336 heat-killed mWasabi-expressing *Mm* (heat-killed by incubation at 80°C for 20 min). To assess succinate levels, larvae were infected with 200 to 300 tdTomato-expressing *Mm*.

TNF and drug administration to zebrafish larvae

TNF^{hi} animals were created by injecting recombinant zebrafish soluble TNF (*42*) as described (*4*). To assess drug treatment in infected animals, equivalently infected sibling larvae were mixed in a petri dish and held at 28.5°C before random allocation to the drug-treated or control groups; 0.5% DMSO (Sigma) was used as the control (vehicle). Drugs dissolved in DMSO or water were kept in small aliquots at −20°C before administration to 1-dpi larvae by adding them to the water (E2/2 medium). Doses used in this work were based on previous studies or pilot experiments, using the minimum effective concentration without deleterious or toxic effects on larvae for the duration of the experiment (see table S2). FCCP [carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone] (50 nM; Cambridge Bioscience) was administered 1.5 hours before MitoTracker Red CM-H_2-Xros injection. TTFA (1 μM; Cambridge Bioscience), atpenin A5 (2.5 nM; Insight Biotechnology), diethyl succinate (500 nM; reagent plus 99%, Sigma), and DEBM (diethyl butyl malonate) (1 μM; Sigma) were administered 2 hours before MitoTracker Red CM-H_2-Xros injection. DM-glutamate (dimethyl glutamate) (60 μM; Cambridge Bioscience) was administered 3 hours before MitoTracker Red CM-H_2-Xros Injection. DNP (2.4-dinitrophenol) (100 nM; Agilent Technologies) was administered 3.5 hours before MitoTracker Red CM-H_2-Xros injection. Rotenone (6.25 nM; Sigma), piericidin A (50 nM; Stratech Scientific), strobilurin B (100 nM; Insight Biotechnology), metformin (20 μM; VWR International), phenformin (20 μM; Sigma), nigericin (5 μM; Sigma), diazoxide (50 nM; Cambridge Bioscience), UK5099 (10 μM; Cambridge Bioscience), and M-pyruvate (methyl pyruvate) (50 nM; Fisher Scientific) were administered 4 hours before MitoTracker Red CM-H_2-Xros injection. DM-malonate (10 μM; Sigma), perhexiline (10 μM; Stratech Scientific), 4-BrCA (4-bromocrotonic acid) (10 μM; Insight Biotechnology), GPNA (10 μM; Cambridge Bioscience), BPTES (5 μM; Cambridge Bioscience), telaglenastat (5 μM; Cambridge Bioscience), and R-162 (1 μM; Cambridge Bioscience) were administered 5 hours before MitoTracker Red CM-H_2-Xros injection. In experiments to assess cording, perhexiline was removed 5 hours after TNF administration, diethyl succinate and DEBM were administered for 10 hours and then removed, and metformin, phenformin, DM-malonate, diazoxide, and telaglenastat were added at 1 dpi and removed at 2 dpi. After drug removal, the larvae were maintained in fresh E2/2 medium for the rest of the experiment. In experiments to assess macrophage numbers, diethyl succinate and DEBM were administered at 2 dpf/1 dpi for 24 hours until macrophage number was assessed at 3 dpf/2 dpi. For experiments quantifying mitochondrial ROS production, drugs were added before MitoTracker Red CM-H_2-Xros injection as indicated above and maintained during imaging.

Synthetic mRNA synthesis and microinjection

The ORF sequence of the alternative oxidase (AOX) from *Ciona intestinalis* was obtained by PCR using as a template the plasmid MAC_C_AOX (Addgene plasmid #111661). The T7 promoter (5′-TAATACGACTCACTATAGG-3′) followed by the zebrafish Kozak sequence 5′-GCCGCCACC-3′ were inserted before the start codon by PCR. mRNA was synthesized using the mMessage mMachine kit (Ambion) and the polyA Tailing kit (Ambion). Approximately 2 to 4 nl of injection solution (*4*) containing AOX mRNA (200 μg/ml) was injected into the yolks of embryos at the one- to two-cell stage.

Morpholino-mediated knockdown of RIP3 and PGAM5

RIP3 e2/i2 splice-blocking (5′-TTTTAGAAATCACCTTGGCATCCAG-3′) and PGAM5 translation-blocking morpholino (5′-AGCGCCCTCCGAAAAGACATGCTTC-3′) (Gene Tools) were diluted to 0.15 mM in injection solution (*4*). Approximately 2 to 4 nl was injected into the yolks of embryos at the one- to two-cell stage.

Heart rate assessment of zebrafish larvae

AOX-expressing 2-dpf larvae were treated with different concentrations of KCN for an hour. Heart rate (beats per minute) was assessed as a readout of cyanide poisoning of complex IV of the electron transport chain (*43*) in the absence of anesthetic using a dissecting microscope.

Zebrafish larva microscopy

Fluorescence microscopy was performed as described (*39*). Mycobacterial cording and macrophage numbers were assessed in the trunk of the larvae using a Nikon Eclipse E600 upright microscope fitted with Nikon Plan Fluor 10× 0.3 NA and Nikon Plan Fluor 20× 0.5 NA objectives. For laser scanning confocal microscopy, anesthetized larvae were embedded in low-melting-point agarose as described (*6*). A Nikon A1R confocal microscope with a Plan Apo 20× 0.75 NA objective was used to generate 35- to 40-mm *z*-stacks consisting of 0.3- to 2-mm optical sections. The galvano scanner was used for all static imaging and for time-lapse imaging of the caudal hematopoietic tissue (CHT, area located between the cloaca and the beginning of the caudal fin). Images were acquired with NIS Elements (Nikon). A heating chamber (Oko-labs) adapted to the microscope was used to maintain temperature at 28.5°C during imaging. Confocal images are pseudocolored to facilitate visualization.

Mitochondrial ROS quantification assay in zebrafish larvae

Mitochondrial ROS production was assayed by fluorescence intensity of MitoTracker Red CM-H_2-Xros, a cell-permeable fluorogenic probe for ROS, which is targeted to the mitochondrion and produces red fluorescence upon oxidation by diverse ROS (Fisher Scientific) (*5*, *6*). *Tg (mpeg1:YFP)*^{w200} larvae were infected at 2 dpf.

For all experiments where TNF^{hi} animals were used, larvae were microinjected at 1 dpi via CV with phosphate-buffered saline (PBS) containing TNF and 50 mM MitoTracker Red CM-H_2-Xros or PBS containing vehicle for TNF and MitoTracker Red CM-H_2-Xros (*6*). For experiments where mROS production was quantified in mycobacterium-infected versus uninfected macrophages, 100 mM MitoTracker Red CM-H_2-Xros was used instead to increase sensitivity of the probe. After administration of MitoTracker Red CM-H_2-Xros (in combination with TNF or alone), larvae were prepared for confocal imaging and maintained at 28.5°C within a heated incubation chamber attached to the confocal microscope. Images of the CHT of each larva were taken starting 30 to 60 min after MitoTracker Red CM-H_2-Xros administration. Mitochondrial ROS production was quantified using maximum projection images as MitoTracker Red CM-H_2-Xros maximum fluorescence intensity per macrophage using NIS-Elements. When not otherwise stated in the figure legend, the mean of maximum MitoTracker Red CM-H_2-Xros fluorescence was quantified only in *Mm*- or *Mtb*-infected macrophages.

Succinate quantification by liquid chromatography–mass spectrometry

Three to six pools of 20 1-dpi larvae per condition per experiment were collected and flash-frozen 30 min after TNF injection, with the time set after injecting 75% of the larvae for each experimental group. Each pool was homogenized in 300 μl of extraction buffer and succinate was quantified as described (*44*). The means and pooled standard deviations of independent experiments were calculated and compared using one-way analysis of variance (ANOVA) with Tukey's post hoc multiple comparisons test.

Quantification of THP-1 cell necrosis

THP-1 cells (ATCC TIB-202) were differentiated into macrophages and infected with single-cell suspensions of mCherry- or tdTomato-expressing *Mtb* mc^{2}6206 Δ*leuD* Δ*panCD* as described (*6*). In THP-1 experiments with added TNF, 1-dpi cells were pre-incubated with 10 nM rotenone, 1 mM metformin, or 0.1% DMSO vehicle control for 1 hour. Human recombinant TNF (Sigma) in a solution of 5% trehalose/PBS (Sigma) was then added to treatment wells as described (*6*). In the experiment with 5 μM mitoparaquat, drug or 0.1% DMSO vehicle control was added 1 day after *Mtb* infection and images acquired after 5 hours of incubation. SYTOX Green Nucleic Acid Stain (Life Technologies) was added to culture medium 30 min before image acquisition. Macrophages were imaged using a Nikon Ti-E inverted microscope fitted with a 20× objective (Nikon, CFI S Plan Fluor 0.45 NA) and two to five arbitrary images per well acquired with NIS Elements (Nikon). Cell necrosis was quantified as described (*6*).

Statistical analysis

The following statistical analyses were performed using Prism 7 (GraphPad): two-way ANOVA or one-way ANOVA with Dunn's or Tukey's post-test and Fisher's exact test. Error bars denote SEM. Post-test P values were defined as follows: $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$ (not significant, $P > 0.05$). Where the n value is given in a figure and not represented graphically in the figure, n represents the number of zebrafish used for each experimental group.

Software used

The following software was used: NIS-Elements for image acquisition in wide-field and confocal microscopy, ImageJ (https://fiji.sc/) for image analysis of macrophage death, GraphPad Prism 7.0 (GraphPad Software) for data graphing and statistical analyses, and CorelDRAW (CorelDRAW Graphics Suite x5) for figure preparation.

REFERENCES AND NOTES

1. J. Keane *et al*., Tuberculosis associated with infliximab, a tumor necrosis factor α-neutralizing agent. *N. Engl. J. Med.* **345**, 1098–1104 (2001). doi: 10.1056/NEJMoa011110; pmid: 11596589
2. M. A. Behr, P. H. Edelstein, L. Ramakrishnan, Is *Mycobacterium tuberculosis* infection life long? *BMJ* **367**, l5770 (2019). doi: 10.1136/bmj.l5770; pmid: 31649096
3. L. J. Whitworth *et al*., Elevated cerebrospinal fluid cytokine levels in tuberculous meningitis predict survival in response to dexamethasone. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2024852118 (2021). doi: 10.1073/pnas.2024852118; pmid: 33658385
4. D. M. Tobin *et al*., Host genotype-specific therapies can optimize the inflammatory response to mycobacterial infections. *Cell* **148**, 434–446 (2012). doi: 10.1016/j.cell.2011.12.023; pmid: 22304914
5. F. J. Roca, L. Ramakrishnan, TNF dually mediates resistance and susceptibility to mycobacteria via mitochondrial reactive oxygen species. *Cell* **153**, 521–534 (2013). doi: 10.1016/j.cell.2013.03.022; pmid: 23582643
6. F. J. Roca, L. J. Whitworth, S. Redmond, A. A. Jones, L. Ramakrishnan, TNF Induces Pathogenic Programmed Macrophage Necrosis in Tuberculosis through a Mitochondrial-Lysosomal-Endoplasmic Reticulum Circuit. *Cell* **178**, 1344–1361.e11 (2019). doi: 10.1016/j.cell.2019.08.004; pmid: 31474371
7. C. J. Cambier, S. Falkow, L. Ramakrishnan, Host evasion and exploitation schemes of Mycobacterium tuberculosis. *Cell* **159**, 1497–1509 (2014). doi: 10.1016/j.cell.2014.11.024; pmid: 25525872
8. H. Clay, H. E. Volkman, L. Ramakrishnan, Tumor necrosis factor signaling mediates resistance to mycobacteria by inhibiting bacterial growth and macrophage death. *Immunity* **29**, 283–294 (2008). doi: 10.1016/j.immuni.2008.06.011; pmid: 18691913
9. D. M. Tobin *et al*., The lta4h locus modulates susceptibility to mycobacterial infection in zebrafish and humans. *Cell* **140**, 717–730 (2010). doi: 10.1016/j.cell.2010.02.013; pmid: 20211140
10. L. Whitworth *et al*., A Bayesian analysis of the association between *Leukotriene A4 Hydrolase* genotype and survival in tuberculous meningitis. *eLife* **10**, e61722 (2021). doi: 10.7554/eLife.61722; pmid: 33416499
11. M. J. Marakalala *et al*., Inflammatory signaling in human tuberculosis granulomas is spatially organized. *Nat. Med.* **22**, 531–538 (2016). doi: 10.1038/nm.4073; pmid: 27043495
12. M. P. Murphy, How mitochondria produce reactive oxygen species. *Biochem. J.* **417**, 1–13 (2009). doi: 10.1042/BJ20081386; pmid: 19061483
13. F. Scialò, D. J. Fernández-Ayala, A. Sanz, Role of Mitochondrial Reverse Electron Transport in ROS Signaling: Potential Roles in Health and Disease. *Front. Physiol.* **8**, 428 (2017). doi: 10.3389/fphys.2017.00428; pmid: 28701960
14. R. Fato *et al*., Differential effects of mitochondrial Complex I inhibitors on production of reactive oxygen species. *Biochim. Biophys. Acta* **1787**, 384–392 (2009). doi: 10.1016/j.bbabio.2008.11.003; pmid: 19059197
15. E. L. Robb *et al*., Control of mitochondrial superoxide production by reverse electron transport at complex I. *J. Biol. Chem.* **293**, 9869–9879 (2018). doi: 10.1074/jbc.RA118.003647; pmid: 29743240
16. H. Nůsková, M. Vrbacký, Z. Drahota, J. Houštěk, Cyanide inhibition and pyruvate-induced recovery of cytochrome c oxidase. *J. Bioenerg. Biomembr.* **42**, 395–403 (2010). doi: 10.1007/s10863-010-9307-6; pmid: 20725851
17. N. Burger *et al*., A sensitive mass spectrometric assay for mitochondrial CoQ pool redox state in vivo. *Free Radic. Biol. Med.* **147**, 37–47 (2020). doi: 10.1016/j.freeradbiomed.2019.11.028; pmid: 31811922
18. E. T. Chouchani *et al*., Ischaemic accumulation of succinate controls reperfusion injury through mitochondrial ROS. *Nature* **515**, 431–435 (2014). doi: 10.1038/nature13909; pmid: 25383517
19. E. L. Mills *et al*., Succinate Dehydrogenase Supports Metabolic Repurposing of Mitochondria to Drive Inflammatory Macrophages. *Cell* **167**, 457–470.e13 (2016). doi: 10.1016/j.cell.2016.08.064; pmid: 27667687
20. D. W. Zhang *et al*., RIP3, an energy metabolism regulator that switches TNF-induced cell death from apoptosis to necrosis. *Science* **325**, 332–336 (2009). doi: 10.1126/science.1172308; pmid: 19498109
21. V. Goossens, G. Stangé, K. Moens, D. Pipeleers, J. Grooten, Regulation of tumor necrosis factor-induced, mitochondria- and reactive oxygen species-dependent cell death by the electron flux through the electron transport chain complex I. *Antioxid. Redox Signal.* **1**, 285–295 (1999). doi: 10.1089/ars.1999.1.3-285; pmid: 11229440
22. www.proteinatlas.org/.
23. A. Bröer, F. Rahimi, S. Bröer, Deletion of Amino Acid Transporter ASCT2 (SLC1A5) Reveals an Essential Role for Transporters SNAT1 (SLC38A1) and SNAT2 (SLC38A2) to Sustain Glutaminolysis in Cancer Cells. *J. Biol. Chem.* **291**, 13194–13205 (2016). doi: 10.1074/jbc.M115.700534; pmid: 27129276
24. www.sanger.ac.uk/tool/basicz/.
25. C. Q. Zhong *et al*., Quantitative phosphoproteomic analysis of RIP3-dependent protein phosphorylation in the course of TNF-induced necroptosis. *Proteomics* **14**, 713–724 (2014). doi: 10.1002/pmic.201300326; pmid: 24453211
26. A. Stepanova, Y. Shurubor, F. Valsecchi, G. Manfredi, A. Galkin, Differential susceptibility of mitochondrial complex II to inhibition by oxaloacetate in brain and heart. *Biochim. Biophys. Acta* **1857**, 1561–1568 (2016). doi: 10.1016/j.bbabio.2016.06.002; pmid: 27287543
27. Z. Wang, H. Jiang, S. Chen, F. Du, X. Wang, The mitochondrial phosphatase PGAM5 functions at the convergence point of multiple necrotic death pathways. *Cell* **148**, 228–243 (2012). doi: 10.1016/j.cell.2011.11.030; pmid: 22265414
28. H. A. Prag *et al*., Ester Prodrugs of Malonate with Enhanced Intracellular Delivery Protect Against Cardiac Ischemia-Reperfusion Injury In Vivo. *Cardiovasc. Drugs Ther.* **36**, 1–13 (2022). doi: 10.1007/s10557-020-07033-6; pmid: 32648168
29. G. Vial, D. Detaille, B. Guigas, Role of Mitochondria in the Mechanism(s) of Action of Metformin. *Front. Endocrinol.* **10**, 294 (2019). doi: 10.3389/fendo.2019.00294; pmid: 31133988
30. A. Stepanova *et al*., Reverse electron transfer results in a loss of flavin from mitochondrial complex I: Potential mechanism for brain ischemia reperfusion injury. *J. Cereb. Blood Flow Metab.* **37**, 3649–3658 (2017). doi: 10.1177/0271678X17730242; pmid: 28914132
31. G. M. Tannahill *et al*., Succinate is an inflammatory signal that induces IL-1β through HIF-1α. *Nature* **496**, 238–242 (2013). doi: 10.1038/nature11986; pmid: 23535595
32. K. Łabuzek *et al*., Quantification of metformin by the HPLC method in brain regions, cerebrospinal fluid and plasma of rats treated with lipopolysaccharide. *Pharmacol. Rep.* **62**, 956–965 (2010). doi: 10.1016/S1734-1140(10)70357-1; pmid: 21098880

33. A. Singhal *et al*., Metformin as adjunct antituberculosis therapy. *Sci. Transl. Med.* **6**, 263ra159 (2014). doi: 10.1126/scitranslmed.3009885; pmid: 25411472
34. L. Tsenova, A. Singhal, Effects of host-directed therapies on the pathology of tuberculosis. *J. Pathol.* **250**, 636–646 (2020). doi: 10.1002/path.5407; pmid: 32108337
35. C. Padmapriyadarsini *et al*., Evaluation of metformin in combination with rifampicin containing antituberculosis therapy in patients with new, smear-positive pulmonary tuberculosis (METRIF): Study protocol for a randomised clinical trial. *BMJ Open* **9**, e024363 (2019). doi: 10.1136/bmjopen-2018-024363; pmid: 30826761
36. N. K. Dutta, M. L. Pinn, P. C. Karakousis, Metformin Adjunctive Therapy Does Not Improve the Sterilizing Activity of the First-Line Antitubercular Regimen in Mice. *Antimicrob. Agents Chemother.* **61**, 00652-17 (2017). https://doi.org/10.1128/aac.00652-17
37. P. Muthuswamy, T. C. Hu, B. Carasso, M. Antonio, N. Dandamudi, Prednisone as adjunctive therapy in the management of pulmonary tuberculosis. Report of 12 cases and review of the literature. *Chest* **107**, 1621–1630 (1995). doi: 10.1378/chest.107.6.1621; pmid: 7781357
38. R. A. Smego, N. Ahmed, A systematic review of the adjunctive use of systemic corticosteroids for pulmonary tuberculosis. *Int. J. Tuberc. Lung Dis.* **7**, 208–213 (2003). pmid: 12661833
39. K. Takaki, J. M. Davis, K. Winglee, L. Ramakrishnan, Evaluation of the pathogenesis and treatment of Mycobacterium marinum infection in zebrafish. *Nat. Protoc.* **8**, 1114–1124 (2013). doi: 10.1038/nprot.2013.068; pmid: 23680983
40. J. M. Mouton *et al*., Comprehensive Characterization of the Attenuated Double Auxotroph *Mycobacterium tuberculosisΔleuDΔpanCD* as an Alternative to H37Rv. *Front. Microbiol.* **10**, 1922 (2019). doi: 10.3389/fmicb.2019.01922; pmid: 31481950
41. M. M. Osman *et al*., The C terminus of the mycobacterium ESX-1 secretion system substrate ESAT-6 is required for phagosomal membrane damage and virulence. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2122161119 (2022). doi: 10.1073/pnas.2122161119; pmid: 35271388
42. F. J. Roca *et al*., Evolution of the inflammatory response in vertebrates: Fish TNF-α is a powerful activator of endothelial cells but hardly activates phagocytes. *J. Immunol.* **181**, 5071–5081 (2008). doi: 10.4049/jimmunol.181.7.5071; pmid: 18802111
43. A. K. Nath *et al*., Chemical and metabolomic screens identify novel biomarkers and antidotes for cyanide exposure. *FASEB J.* **27**, 1928–1938 (2013). doi: 10.1096/fj.12-225037; pmid: 23345455
44. H. A. Prag *et al*., Mechanism of succinate efflux upon reperfusion of the ischaemic heart. *Cardiovasc. Res.* **117**, 1188–1201 (2021). doi: 10.1093/cvr/cvaa148; pmid: 32766828

ACKNOWLEDGMENTS

We thank J. Walker for sharing his knowledge and insights, advice, critical appraisal of the work through the years, and critical review of the paper; P. Edelstein for help and advice on statistical analysis and critical review of the paper; D. Tobin and E. Kunji for critical review of the paper; N. Burger for discussion on mass spectrometry assays and results; J. Baeck for toxicity testing of several of the compounds used and preliminary assessments of the effects of some; K. Takaki, J. Fan, and B. Lyu for help with experiments to quantify succinate; N. Goodwin, R. Foster, and the University of Cambridge aquatics facility staff for zebrafish husbandry; and the LMB's media service for preparation of bacterial and tissue culture reagents. **Funding:** Supported by Wellcome Trust Principal Research Fellowship 223103/Z/21/Z and NIH MERIT award R37 AI054503 (L.R.), Medical Research Council UK grant MC_U105663142 and Wellcome Trust Investigator award 110159/A/15/Z (M.P.M.), and MCIN and "ESF Investing in your Future" Fellowship RYC2019-027799-I/AEI/10.13039/501100011033 (F.J.R.). **Author contributions:** F.J.R. and L.R. designed the project. F.J.R., L.J.W., and H.A.P. performed experiments. F.J.R., L.J.W., H.A.P., L.R., and M.P.M. designed experiments and analyzed and interpreted data. F.J.R. and L.R. wrote the paper. All authors edited the paper with input from the other authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** To facilitate open access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. This work is licensed under a Creative Commons Attribution 4.0 International License.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abh2841
Figs. S1 to S5
Tables S1 and S2
References (*45*–*64*)
Data S1
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 26 February 2021; resubmitted 24 November 2021
Accepted 9 May 2022
10.1126/science.abh2841

RESEARCH ARTICLE SUMMARY

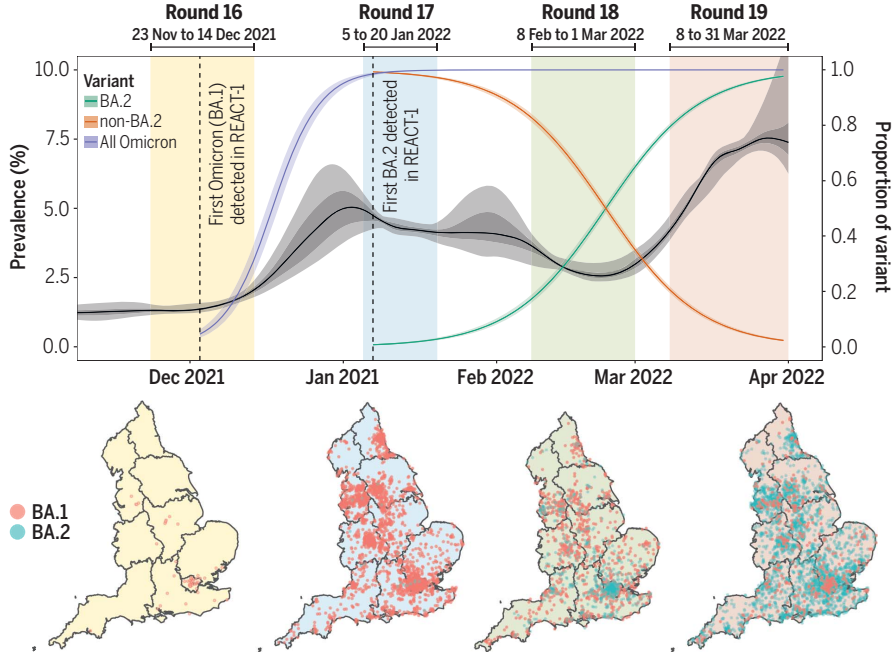
CORONAVIRUS

Twin peaks: The Omicron SARS-CoV-2 BA.1 and BA.2 epidemics in England

Paul Elliott*, Oliver Eales†, Nicholas Steyn†, David Tang†, Barbara Bodinier, Haowei Wang, Joshua Elliott, Matthew Whitaker, Christina Atchison, Peter J. Diggle, Andrew J. Page, Alexander J. Trotter, Deborah Ashby, Wendy Barclay, Graham Taylor, Helen Ward, Ara Darzi, Graham S. Cooke, Christl A. Donnelly*‡, Marc Chadeau-Hyam*‡

BACKGROUND: Rapid transmission of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Omicron variant has led to record-breaking COVID-19 incidence rates around the world. Omicron sublineages have been detected in many countries, with BA.1 replacing Delta and BA.2 replacing BA.1.

RATIONALE: We report here findings from the Real-time Assessment of Community Transmission-1 (REACT-1) study, which has tracked the spread of SARS-CoV-2 in England through 19 distinct rounds of data collection from 1 May 2020 to 31 March 2022. In each round, randomly selected participants aged 5 years and older provided self-administered throat and nose swabs, which were tested using reverse transcription polymerase chain reaction (RT-PCR). The use of random samples rather than reliance on routine testing of symptomatic individuals means that prevalence estimates are not influenced by test-seeking behaviors or availability of tests and include asymptomatic as well as symptomatic infections. With large numbers of individuals at each round (~100,000 or more) and questionnaire data on individual demographics, symptoms, household composition, occupation, lifestyle, and other factors, along with the RT-PCR results, we have been able to obtain high-resolution real-time data on the distribution of SARS-CoV-2 infection in England by time, person, and place. In total, 2,512,797 individuals with a valid RT-PCR test result took part from among 14,036,117 who were sent invitation letters across the 19 rounds. Here, we focus on the Omicron wave spanning round 16 (23 November to 14 December 2021, N = 97,089), round 17 (5 to 20 January 2022, N = 102,174), round 18 (8 February to 1 March 2022, N = 94,950), and round 19 (8 to 31 March 2022, N = 109,181). Viral genome sequencing of positive samples was conducted where cycle threshold values were ≤34 for either the E or the N gene.



"Twin peaks" in SARS-CoV-2 prevalence in England in January and March 2022. (Top) Smoothed daily weighted prevalence from a penalized spline (P-spline) model (black curve, left *y* axis) for January and March 2022, when Omicron BA.1 (blue curve, right *y* axis) and, subsequently, BA.2 (green curve, right *y* axis) were dominant. Solid lines represent posterior estimates of the models, whereas light-shaded regions show the 95% credible intervals of the model estimates. For smoothed prevalence only, the 50% credible interval is also represented as a darker-shaded region. Yellow-, blue-, green-, and orange-shaded regions represent the swab collection days for rounds 16 to 19, respectively. (Bottom) Geographical distribution of Omicron swab positivity by round, showing transition from the first Omicron (BA.1) peak to the second (BA.2) peak.

RESULTS: In March 2022, weighted prevalence was 6.37% (N = 109,181), the highest seen in REACT-1 since the start of data collection in May 2020 and more than twice that in February 2022 following the initial Omicron (BA.1) peak in January 2022. The highest weighted prevalence was seen in children aged 5 to 11 years (almost all unvaccinated) at 8.81% [95% credible interval (CrI): 7.89%, 9.82%], nearly double that seen in older children aged 12 to 17 years at 4.71% (95% CrI: 4.21%, 5.25%), more than 70% of whom had been vaccinated. Prevalence during March 2022 was increasing overall [reproduction number (R) = 1.07; 95% CrI: 1.06, 1.09], with the greatest increase in those aged 65 to 74 years and 75 years and older (R = 1.13; 95% CrIs, respectively: 1.09, 1.17 and 1.08, 1.19). Whereas rates in these older groups were still rising throughout March 2022, they began to fall from their peak in the 5 to 11 year olds (R = 0.94; 95% CrI: 0.90, 0.98). Of the lineages circulating in March 2022 that were determined by viral genome sequencing, 4444 (99.98%) were Omicron—including 405 (9.11%) BA.1, 4026 (90.6%) BA.2, and 13 (0.16%) BA.3 sublineage—and one (0.02%) was Delta AY.4. BA.2 had a daily growth rate advantage (compared with BA.1 and its sublineages) of 0.10 (95% CrI: 0.10, 0.11).

CONCLUSION: The almost complete replacement of Omicron BA.1 by BA.2 resulted in a second peak in SARS-CoV-2 infections and unprecedented levels of infection in England in March 2022, after the initial Omicron (BA.1) peak in January 2022. These high infection rates were associated with increases in hospitalizations and deaths but at much lower levels than in previous waves against a backdrop of high levels of vaccination. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: p.elliott@imperial.ac.uk (P.E.); c.donnelly@imperial.ac.uk (C.A.D.); m.chadeau@imperial.ac.uk (M.C.-H.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
This is an open-access article distributed under the terms of the Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Cite this article as P. Elliott *et al.*, *Science* **376**, eabq4411 (2022). DOI: 10.1126/science.abq4411

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abq4411

RESEARCH ARTICLE

CORONAVIRUS

Twin peaks: The Omicron SARS-CoV-2 BA.1 and BA.2 epidemics in England

Paul Elliott[1,2,3,4,5,6]*, Oliver Eales[1,7]†, Nicholas Steyn[8,9]†, David Tang[1,2]†, Barbara Bodinier[1,2], Haowei Wang[1,7], Joshua Elliott[3,10], Matthew Whitaker[1,2], Christina Atchison[1,3], Peter J. Diggle[11], Andrew J. Page[12], Alexander J. Trotter[12], Deborah Ashby[1], Wendy Barclay[10], Graham Taylor[10], Helen Ward[1,3,4,7], Ara Darzi[3,4,13], Graham S. Cooke[3,4,10], Christl A. Donnelly[1,7,8]*‡, Marc Chadeau-Hyam[1,2]*‡

Rapid transmission of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Omicron variant has led to record-breaking incidence rates around the world. The Real-time Assessment of Community Transmission-1 (REACT-1) study has tracked SARS-CoV-2 infection in England using reverse transcription polymerase chain reaction (RT-PCR) results from self-administered throat and nose swabs from randomly selected participants aged 5 years and older approximately monthly from May 2020 to March 2022. Weighted prevalence in March 2022 was the highest recorded in REACT-1 at 6.37% (*N* = 109,181), with the Omicron BA.2 variant largely replacing the BA.1 variant. Prevalence was increasing overall, with the greatest increase in those aged 65 to 74 years and 75 years and older. This was associated with increased hospitalizations and deaths, but at much lower levels than in previous waves against a backdrop of high levels of vaccination.

Since the emergence of Omicron as the dominant severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variant in England in mid- to late-December 2021 (*1*, *2*), the peak in January 2022 that was associated with the BA.1 variant was the highest prevalence recorded in England up to that time (*3*). This was followed by the replacement of BA.1 by the more transmissible BA.2 variant (*3*–*5*). By late March 2022, BA.2 infections had surged in many European countries (*6*) and had become the dominant variant in the United States (*7*). Although Omicron infections led to fewer severe outcomes than Delta infections, the risk reduction depends on age (*8*) and background immunity, with some deaths observed in children under 12 years of age. In less-vaccinated populations such as Hong Kong and Shanghai that have limited protection from vaccination among their oldest citizens (*9*), BA.2 has spread very quickly (*10*, *11*); the Omicron wave has caused more than 97% of the total COVID-19 death toll in Hong Kong to date (with 9146 deaths between 31 December 2021 and 13 May 2022 out of a total of 9359 COVID-19–related deaths between 31 December 2019 and 13 May 2022) (*12*, *13*).

In England, during the first phase of the Omicron (BA.1) epidemic (*1*), the country saw peaks in hospital admissions in late December 2021 to early January 2022 and in deaths (within 28 days of a positive test) in mid-January 2022 (*14*). Since then, after decreases in February 2022, hospital admissions in late March 2022 returned to levels similar to those seen in January 2022 (~2000 admissions per day on average), with deaths also increasing in England since early March 2022.

The Real-time Assessment of Community Transmission-1 (REACT-1) study has tracked the spread of the SARS-CoV-2 virus among randomly selected community samples in England since May 2020 (*15*). Unlike relying on testing of symptomatic individuals to estimate prevalence, as is the case in most countries, the use of random samples of the population means that estimates are unbiased with respect to test-seeking behaviors and availability of tests and include asymptomatic as well as symptomatic infections (*16*). With the completion of the 19th and final round of REACT-1 data collection, we document here the transmission dynamics of SARS-CoV-2 in England since May 2020 and, in particular, the emergence of the Omicron epidemic and the replacement of BA.1 and its sublineages by BA.2.

[1]School of Public Health, Imperial College London, London, UK. [2]MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [3]Imperial College Healthcare NHS Trust, London, UK. [4]National Institute for Health Research Imperial Biomedical Research Centre, London, UK. [5]Health Data Research (HDR) UK, Imperial College London, London, UK. [6]UK Dementia Research Institute, Imperial College London, London, UK. [7]MRC Centre for Global infectious Disease Analysis and Jameel Institute, Imperial College London, London, UK. [8]Department of Statistics, University of Oxford, Oxford, UK. [9]Department of Mathematics, Imperial College London, London, UK. [10]Department of Infectious Disease, Imperial College London, London, UK. [11]CHICAS, Lancaster Medical School, Lancaster University, UK, and Health Data Research, Lancaster, UK. [12]Quadram Institute, Norwich, UK. [13]Institute of Global Health Innovation, Imperial College London, London, UK.

***Corresponding author. Email: p.elliott@imperial.ac.uk (P.E.); c.donnelly@imperial.ac.uk (C.A.D.); m.chadeau@imperial.ac.uk (M.C.-H.)**

†These authors contributed equally to this work.

‡These authors contributed equally to this work.

Overall prevalence and temporal trends

Of 697,055 individuals invited into the final (19th) round of REACT-1, 109,181 (15.7%) registered and returned a throat and nasal swab (from 8 to 31 March 2022) with a valid SARS-CoV-2 reverse transcription polymerase chain reaction (RT-PCR) test result. Of these, 6902 were positive, yielding a weighted prevalence of 6.37% [95% credible interval (CrI): 6.21%, 6.53%], which is the highest weighted prevalence observed throughout the REACT-1 study (Table 1).

Prevalence levels by demographic and other characteristics are shown in table S1, A and B. Weighted prevalence during round 19 was 7.55% (95% CrI: 7.22%, 7.89%) in households with one or more children compared with 5.89% (95% CrI: 5.71%, 6.08%) in households without children; 17.8% (95% CrI: 17.2%, 18.5%) among those reporting contact with a confirmed COVID-19 case compared with 4.00% (95% CrI: 3.86%, 4.16%) among those without such contact; and 27.6% (95% CrI: 26.7%, 28.5%) among those reporting "classic" COVID-19 symptoms (loss or change of sense of smell or taste, fever, new persistent cough) compared with 2.60% (95% CrI: 2.47%, 2.74%) among those without symptoms. Our results did not show substantial differences in weighted prevalence in relation to smoking or vaping status.

Multivariable logistic regression models showed an increased risk of swab positivity in those living (i) in larger households that include three to five persons (versus single-person households) with a mutually adjusted odds ratio (OR) of 1.09 [95% confidence interval (CI): 1.01, 1.17] and (ii) with one or more children (versus households without children) with a mutually adjusted OR of 1.09 (95% CI: 1.01, 1.18) (table S2). Results also showed a lower risk of swab positivity in those living (i) in urban (versus rural) (*17*) areas with a mutually adjusted OR of 0.93 (95% CI: 0.88, 0.99) and (ii) in deprived [index of multiple deprivation (*18*) in the first and second quintiles] versus most-affluent (fifth quintile) areas with mutually adjusted ORs of 0.82 (95% CI: 0.75, 0.90) and 0.91 (95% CI: 0.84, 0.98), respectively. This last finding is in contrast to our results in earlier rounds that showed higher infection prevalence in more-deprived areas (*19*) but is consistent with a report from the Office for National Statistics Coronavirus Infection Survey that covered the 2 weeks up to 23 April 2022 (when BA.2 dominated), which also found higher infection prevalence in less-deprived areas (*20*).

Trends in prevalence, growth rate, predominant variants, lockdown periods in England, and mobility data over the 23 months from

Table 1. Unweighted and weighted prevalence of SARS-CoV-2 swab positivity from REACT-1 across rounds 1 to 19.

Round	Tested swabs	Positive swabs	Unweighted prevalence (95% CI)	Weighted prevalence (95% CrI)	Date of first sample (DD/MM/YY)	Date of last sample (DD/MM/YY)
1	120,620	159	0.13% (0.11%, 0.15%)	0.16% (0.13%, 0.19%)	01/05/20	01/06/20
2	159,199	123	0.08% (0.07%, 0.09%)	0.09% (0.07%, 0.11%)	19/06/20	07/07/20
3	162,821	54	0.03% (0.03%, 0.04%)	0.04% (0.03%, 0.05%)	24/07/20	11/08/20
4	154,325	137	0.09% (0.08%, 0.11%)	0.13% (0.01%, 0.15%)	20/08/20	08/09/20
5	174,949	824	0.47% (0.44%, 0.50%)	0.60% (0.55%, 0.71%)	18/09/20	05/10/20
6	160,175	1732	1.08% (1.03%, 1.13%)	1.30% (1.21%, 1.39%)	16/10/20	02/11/20
7	168,181	1299	0.77% (0.73%, 0.82%)	0.94% (0.87%, 1.01%)	13/11/20	03/12/20
8	167,642	2282	1.36% (1.31%, 1.42%)	1.57% (1.49%, 1.66%)	06/01/21	22/01/21
9	165,456	689	0.42% (0.39%, 0.45%)	0.49% (0.44%, 0.55%)	04/02/21	23/02/21
10	140,844	227	0.16% (0.14%, 0.18%)	0.20% (0.17%, 0.23%)	11/03/21	30/03/21
11	127,408	115	0.09% (0.07%, 0.11%)	0.10% (0.08%, 0.13%)	15/04/21	03/05/21
12*	108,911	135	0.12% (0.10%, 0.15%)	0.15% (0.12%, 0.18%)	20/05/21	07/06/21
13	98,233	527	0.54% (0.49%, 0.58%)	0.63% (0.57%, 0.69%)	24/06/21	12/07/21
14†	100,527	764	0.76% (0.71%, 0.82%)	0.83% (0.76%, 0.89%)	09/09/21	27/09/21
15‡	100,112	1399	1.40% (1.33%, 1.47%)	1.57% (1.48%, 1.66%)	19/10/21	05/11/21
16§	97,089	1192	1.23% (1.16%, 1.30%)	1.41% (1.33%, 1.51%)	23/11/21	14/12/21
17¶	102,174	4073	3.99% (3.87%, 4.11%)	4.41% (4.25%, 4.56%)	05/01/22	20/01/22
18#**	94,950	2731	2.88% (2.77%, 2.98%)	2.88% (2.76%, 3.00%)	08/02/22	01/03/22
19**	109,181	6902	6.32% (6.18%, 6.47%)	6.37% (6.21%, 6.53%)	08/03/22	31/03/22

*The sampling strategy changed for round 12 and subsequent rounds. Therefore, unweighted prevalence is not directly comparable with that of previous rounds. †N = 509 samples from 28 to 30 September 2021 are included. Sample handling changed in round 14. Therefore, prevalence is not directly comparable with that of previous rounds. ‡N = 93 samples (all negatives) from 6 to 8 November 2021 and N = 86 samples with no collection or arrival dates are included. §N = 661 samples (including 12 positives) from 15 to 17 December 2021 are included. Swab positivity was assessed using a multiplex assay from round 16 onward. Test diagnostic characteristics may differ slightly from those of previous rounds. ¶N = 862 (including 36 positives) from 21 to 24 January 2022 are included. #N = 685 (including 18 positives) from 2 to 4 March 2022 are included. **Incentives were used in rounds 18 and 19 to increase the response rates in previously underrepresented groups.

1 May 2020 to 31 March 2022 are shown in Fig. 1, A to E. A penalized-spline (P-spline) model fit to all REACT-1 data shows an initial decline during the first lockdown in England and increases in the second wave from autumn 2020 through January 2021, the Delta wave in summer to November 2021, and the initial Omicron wave from December 2021 to January 2022. We observed a fall in weighted prevalence (though still at high levels) during February 2022 followed by a steep increase in March 2022 (Fig. 1, A and B). We estimated a doubling time in weighted prevalence of 30.5 (95% CrI: 25.8, 37.0) days in round 19 (8 to 31 March 2022), corresponding to a within-round reproduction number (R) of 1.07 (95% CrI: 1.06, 1.09) with a >0.99 posterior probability that $R > 1$ (Table 2).

From the estimated daily growth rates (Fig. 1C), we clearly show periods of rapid growth associated with (i) the second wave in England as the Alpha variant replaced the wild type in late summer and autumn 2020, (ii) Delta replacing Alpha in late spring and summer 2021, (iii) Omicron replacing Delta in November to December 2021, and (iv) BA.2 replacing BA.1 during February and March 2022 (Fig. 1, E and F). The growth rate plateaued at 0.06 (95% CrI: 0.03, 0.07) on or about 10 March 2022 (Fig. 1C). March 2022 also corresponded to a period of high and increasing mobility (Fig. 1D); by 31 March, indices for driving, walking, and transit reached 92.9, 85.0, and 84.2%, respectively, of the maximum observed throughout the study period from May 2020.

During the Omicron wave, we observed peaks in weighted prevalence during round 17 (5 to 20 January 2022) and round 19, with the highest prevalence in both rounds in individuals aged 5 to 11 years (Fig. 2A and table S1A). P-splines fit to daily weighted prevalence in nine age groups (Fig. 2B) indicated a steep increase in weighted prevalence between round 18 (8 February to 1 March 2022) and round 19 at all ages, followed by (i) a peak in those aged 5 to 11 years around 16 or 17 March 2022, with a subsequent fall; (ii) a monotonic increase throughout round 19 in those aged 12 to 17, 18 to 24, 55 to 64, 65 to 74, and 75+ years; and (iii) an indication that the weighted prevalence may have peaked by the end of March 2022, which was subsequently shown to be the case in the Office for National Statistics Coronavirus Infection Survey data (*21*). Exponential models fit to data from round 19 in the nine age groups showed an overall within-round decreasing prevalence in those aged 5 to 11 years with a within-round R of 0.94 (95% CrI: 0.90, 0.98) and <0.01 posterior probability that $R > 1$, whereas there was a within-round increasing prevalence at all other ages (with ≥0.99 posterior probability that $R > 1$) except for those aged 45 to 54 years (Table 2).

In round 13 (24 June to 12 July 2021) (*22*), only 2.86% (weighted estimate) of children aged 12 to 17 years had been vaccinated. At that time, the weighted prevalence of SARS-CoV-2 swab positivity was 1.53 (95% CrI: 1.00, 2.06) times higher among those aged 12 to 17 years than among those aged 5 to 11 years. Since then, as the vaccine program in older children in England took off, the ratio of weighted prevalence in children aged 12 to 17 years relative to that of those aged 5 to 11 years (almost all unvaccinated) dropped to 0.53 (95% CrI: 0.28, 0.78) in round 19.

Geographic trends

The region-specific weighted prevalence in round 19 ranged from 5.28% (95% CrI: 4.85%, 5.75%) in West Midlands to 8.13% (95% CrI: 7.59%, 8.71%) in South West (Fig. 3A and table S1A) with a within–round 19 $R > 1$ in all regions except London (Table 2). Nearest-neighbor smoothed estimates (see Materials and methods) indicated "twin peaks" in weighted prevalence in rounds 17 and 19. There was a strong north-to-south decreasing prevalence gradient in round 17 (Fig. 3C) but a strong south-to-north decreasing gradient in round 19 (Fig. 3E) with

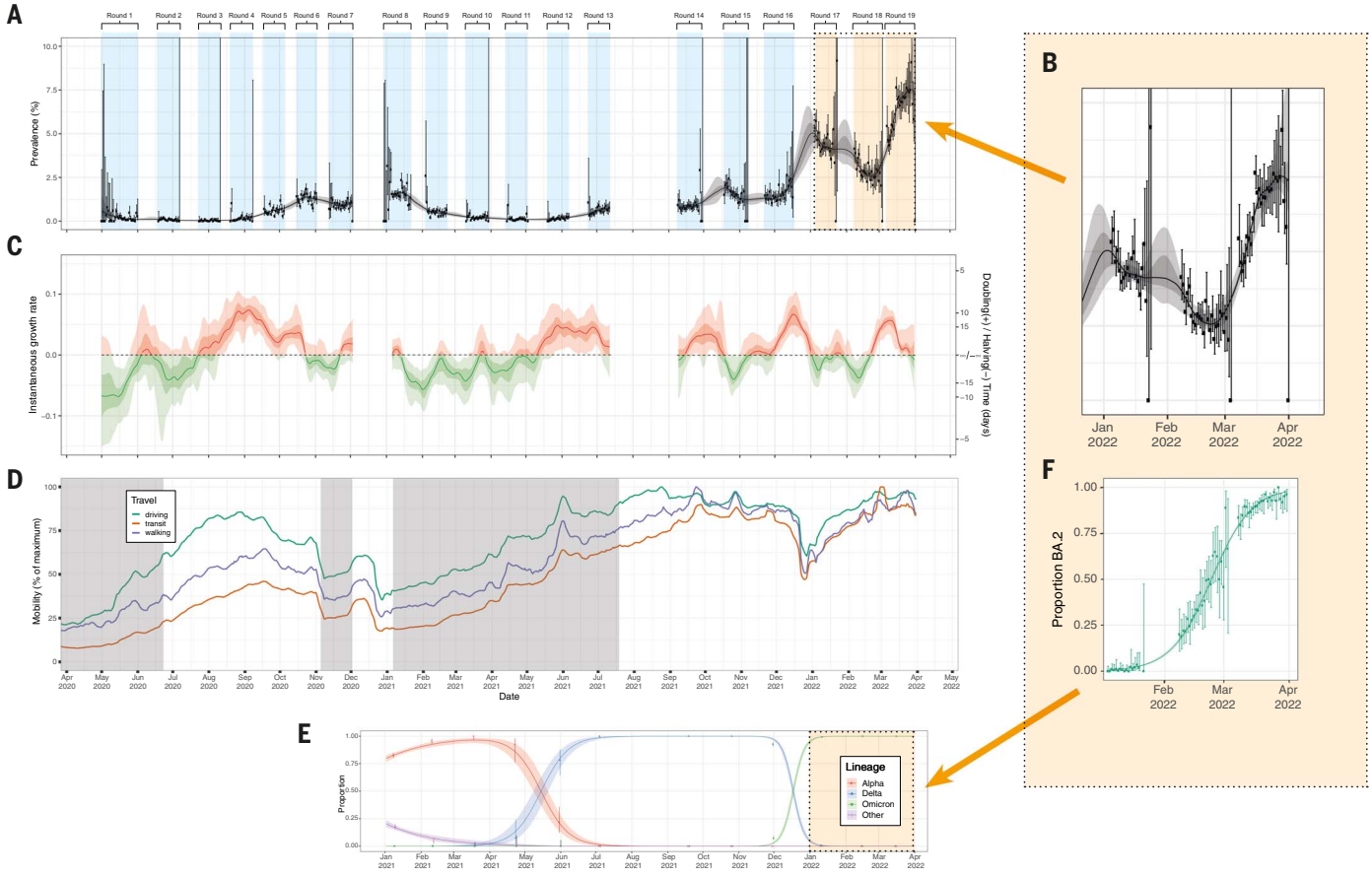


Fig. 1. Overview of SARS-CoV-2 swab positivity across the 19 rounds of the REACT-1 study. (**A**) P-spline model fit to all rounds of REACT-1. The gray-shaded region shows 50% (dark gray) and 95% (light gray) posterior CrI for the P-spline model. Weighted prevalence of swab positivity (*y* axis) is represented for each day of sampling (*x* axis). Weighted observations (black dots) and 95% CrIs (vertical lines) are also shown on an ordinal scale. (**B**) Magnified view of the P-spline model for rounds 16 to 19. (**C**) Instantaneous growth rate for each of the swab days of the REACT study from the log-transformed P-spline. Posterior median estimates are represented as a smoothed solid line, and 50% (dark-shaded regions) and 95% (light-shaded regions) are plotted in red for positive and in green for negative growth rates. (**D**) Daily Apple mobility indices for walking, driving, and transit from phone location data for the duration of the REACT-1 study (1 May 2020 to 31 March 2022). We report 7-day moving averages of the indices and have scaled them to the maximum observed during the study period. Gray-shaded regions represent periods when lockdown was implemented in England. (**E**) Daily proportion of Alpha (red), Delta (blue), Omicron (green), and other (purple) SARS-CoV-2 lineages across rounds 8 to 19 of the REACT-1 study. Mean daily proportions (solid lines) and their 95% CrIs (shaded regions) are shown. (**F**) Daily proportion of infections with BA.2 and its sublineages (versus all other Omicron sublineages) among positive swabs with determined lineage with at least 50% genome coverage in rounds 17 to 19. Point estimates are represented (dots) along with 95% CIs (vertical lines). Smoothed estimates of the proportion are also shown (solid line) together with their 95% CrIs (shaded regions).

an estimated smoothed prevalence >8.0% for 19 lower-tier local authorities (LTLAs), all in South West and East of England, consistent with our findings of higher rates in rural compared with urban areas.

Viral genome sequencing

Viral genome sequencing of the 5621 positive samples obtained in round 19 resulted in 4445 (79.1%) determined lineages with more than 50% genome coverage; one (0.02%; 95% CI: 0.00%, 0.13%) was the AY.4 Delta sublineage, whereas all others were Omicron sublineages (table S3). Among these, 9.11% (95% CI: 8.28%, 10.0%; N = 405) corresponded to BA.1 or its sublineages, 90.6% (95% CI: 89.7%, 91.4%; N = 4026) to BA.2 or its sublineages, and 0.29% (95% CI: 0.16%, 0.50%; N = 13) to BA.3. Ten samples with >90% genome coverage were identified as BA.1 and BA.2 recombinants (N = 7 XE, N = 2 XL, and N = 1 XJ).

Using exponential models, we estimated a daily growth-rate advantage of 0.10 (95% CrI: 0.10, 0.11) in the odds of an infection being from BA.2 (versus all other Omicron sublineages), a 97.7% (95% CI: 97.4%, 98.0%) proportion of BA.2 as of 31 March 2022 (Fig. 1F), and an estimated 56.4 (95% CrI: 54.2, 58.7) days for the proportion of BA.2 to grow from 5 to 95%. This is an approximately twofold lower rate than our estimate for the Delta-to-Omicron transition (28.5 days; 95% CI: 26.3, 30.7), more than 20% higher than the Alpha-to-Delta transition, and almost fourfold higher than the wild type–to-Alpha transition (Table 3).

Discussion

Within a purpose-designed series of large cross-sectional population-based surveys with a random selection of participants, we document here the transmission dynamics of SARS-CoV-2 in England from 1 May 2020 to 31 March 2022, at the height of the Omicron BA.2 wave. The Omicron epidemic in England was characterized by two distinct phases: (i) very rapid replacement of Delta by Omicron during December 2021 and early January 2022, leading to the highest rates of infection since the start of REACT-1; and (ii) rapid replacement of Omicron BA.1 and its sublineages by BA.2 during February to March 2022. Mobility indices also reached their highest levels (since October 2021) in March 2022, reflecting increased social mixing as restrictions were

Table 2. Growth rates per day, reproduction numbers, and doubling or halving times (in days) of SARS-CoV-2 swab positivity from exponential model fits on data from round 19 (8 to 31 March 2022). Numbers in parentheses indicate CrIs.

Category	Group	Growth rate per day (*r*)	Reproduction number (*R*)*	Probability *R* > 1, *r* > 0	Doubling (+) or halving (−) time (days)
All positives		0.023 (0.019, 0.027)	1.07 (1.06, 1.09)	>0.99	30.5 (37.0, 25.8)
Age	Aged 5 to 11	−0.017 (−0.029, −0.005)	0.94 (0.90, 0.98)	<0.01	−41.0 (−23.9, †)
	Aged 12 to 17	0.021 (0.003, 0.039)	1.07 (1.01, 1.13)	0.99	33.1 (†, 17.8)
	Aged 18 to 24	0.028 (0.013, 0.043)	1.09 (1.04, 1.14)	>0.99	24.6 (†, 16.2)
	Aged 25 to 34	0.028 (0.018, 0.038)	1.09 (1.06, 1.12)	>0.99	24.7 (38.7, 18.2)
	Aged 35 to 44	0.018 (0.007, 0.028)	1.06 (1.02, 1.09)	>0.99	39.4 (†, 24.8)
	Aged 45 to 54	0.008 (−0.002, 0.019)	1.03 (0.99, 1.06)	0.94	† (†, 37.2)
	Aged 55 to 64	0.029 (0.018, 0.041)	1.10 (1.06, 1.13)	>0.99	23.6 (38.8, 17.0)
	Aged 65 to 74	0.040 (0.027, 0.053)	1.13 (1.09, 1.17)	>0.99	17.3 (25.6, 13.1)
	Aged 75 and older	0.041 (0.024, 0.057)	1.13 (1.08, 1.19)	>0.99	17.1 (29.0, 12.2)
Region	East Midlands	0.019 (0.004, 0.033)	1.06 (1.01, 1.11)	0.99	37.1 (†, 21.0)
	West Midlands	0.022 (0.008, 0.036)	1.07 (1.03, 1.12)	>0.99	31.5 (†, 19.4)
	East of England	0.034 (0.022, 0.045)	1.11 (1.07, 1.15)	>0.99	20.7 (31.6, 15.5)
	London	0.004 (−0.007, 0.014)	1.01 (0.98, 1.05)	0.75	† (†, 49.7)
	North West	0.029 (0.017, 0.041)	1.09 (1.05, 1.13)	>0.99	24.2 (41.7, 17.1)
	North East	0.041 (0.020, 0.062)	1.14 (1.07, 1.20)	>0.99	16.8 (34.6, 11.2)
	South East	0.016 (0.006, 0.026)	1.05 (1.02, 1.08)	>0.99	42.9 (†, 27.0)
	South West	0.025 (0.013, 0.036)	1.08 (1.04, 1.12)	>0.99	28.3 (†, 19.4)
	Yorkshire and The Humber	0.045 (0.031, 0.059)	1.15 (1.10, 1.19)	>0.99	15.5 (22.1, 11.8)

*Within-round reproduction number (*R*) was calculated assuming an Omicron-specific gamma-distributed generation time, with mean 3.3 days and standard deviation of 3.5 days. †Doubling or halving times had an estimated magnitude greater than 50 days and thus represented approximately constant prevalence.

eased. Weighted prevalence in March 2022 increased most rapidly among older adults, who, despite high levels of vaccination, remain the most vulnerable to serious illness, hospitalizations, and death from COVID-19. Hospitalizations and deaths from COVID-19 in England also increased in March 2022 as infections were rising (*14*), but at much lower rates than in previous waves, reflecting the high levels of vaccination in the population (*14*).

We show a transmission advantage for Alpha compared with wild type in the second wave of infections in England, peaking in January 2021; for Delta as it replaced Alpha during April to June 2021; for Omicron (BA.1) as it rapidly replaced Delta; and, most recently, for BA.2 versus BA.1. Data showing a transmission advantage for BA.2 compared with BA.1 have also been reported from the national routine testing data in the UK (*23*) and in Denmark (*24*).

The Omicron epidemic in England, which involved twin peaks as the epidemic transitioned from Delta to Omicron (BA.1) and then from BA.1 to BA.2, has unfolded over a 3-month period, ahead of similar epidemics in most other countries. Although the immune landscape differs by country and over time because of both natural infection and vaccination (*16*), the transmission dynamics in England are highly relevant to other high-income countries that, like England, experienced Alpha, Delta, and, subsequently, Omicron waves of infection alongside an extensive vaccination program.

The transmission advantage for one variant over another will depend on the immune background (which varies over time) as well as transmissibility (*25*) and mean generation time, so it is not possible to assess directly how much more intrinsically transmissible Omicron is compared with wild type or Alpha; nonetheless, our data show that each new variant of concern has demonstrated a transmission advantage over the previous variants. In addition, we detected recombinant infections, notably XE (BA.1 and BA.2 recombinant). Little is known about the clinical manifestations of XE and whether it may lead to more severe disease than BA.1 or BA.2, but early indications suggest a growth advantage compared with BA.2 (*26*). Continued surveillance of such recombinant infections is warranted.

To deal with the initial phase of the Omicron epidemic, some countries reinitiated social distancing policies (*27*), whereas, in the United States, health care systems struggled to cope with the increase in health care demands (*28*) and, in England, the vaccination program was accelerated. Subsequently, the UK government (on 24 February 2022) removed all domestic legal restrictions concerning COVID-19 in England (*29*) as part of the government's plan for "Living with COVID-19" (*30*); as of 24 February 2022, the legal requirement to self-isolate for COVID-19 was lifted, and since 1 April 2022, all remaining restrictions in England have been removed (*29*). Also, as of 1 April 2022, with a few exceptions, free lateral flow and PCR tests were no longer available and other surveillance measures were curtailed, with greater reliance on the vaccine program to manage the ongoing epidemic. In this regard, our most recent data on infections in children show, through most of March 2022, much higher infection rates in 5 to 11 year olds (for whom vaccination rollout only commenced April 2022) than in 12 to 17 year olds, more than 70% of whom had been vaccinated (one or two doses) by the end of March 2022 (*31*). The drop in the prevalence in swab positivity among 5 to 11 year olds from mid-March 2022 from very high levels over a prolonged period could suggest some depletion, at least temporarily, in their population-level susceptibility due to high levels of natural infection.

Our study has limitations. We rely on unsupervised self-swabbing at home by named individuals selected at random from the National Health Service (NHS) registers. Although response rates of more than 30% were achieved during the first lockdown in England in May 2020, they fell to 12.2% by round 17 (January 2022). We included a small monetary incentive in rounds 18 and 19 (February to March 2022) among participants aged 13 to 44 years, which increased overall response rates in those rounds

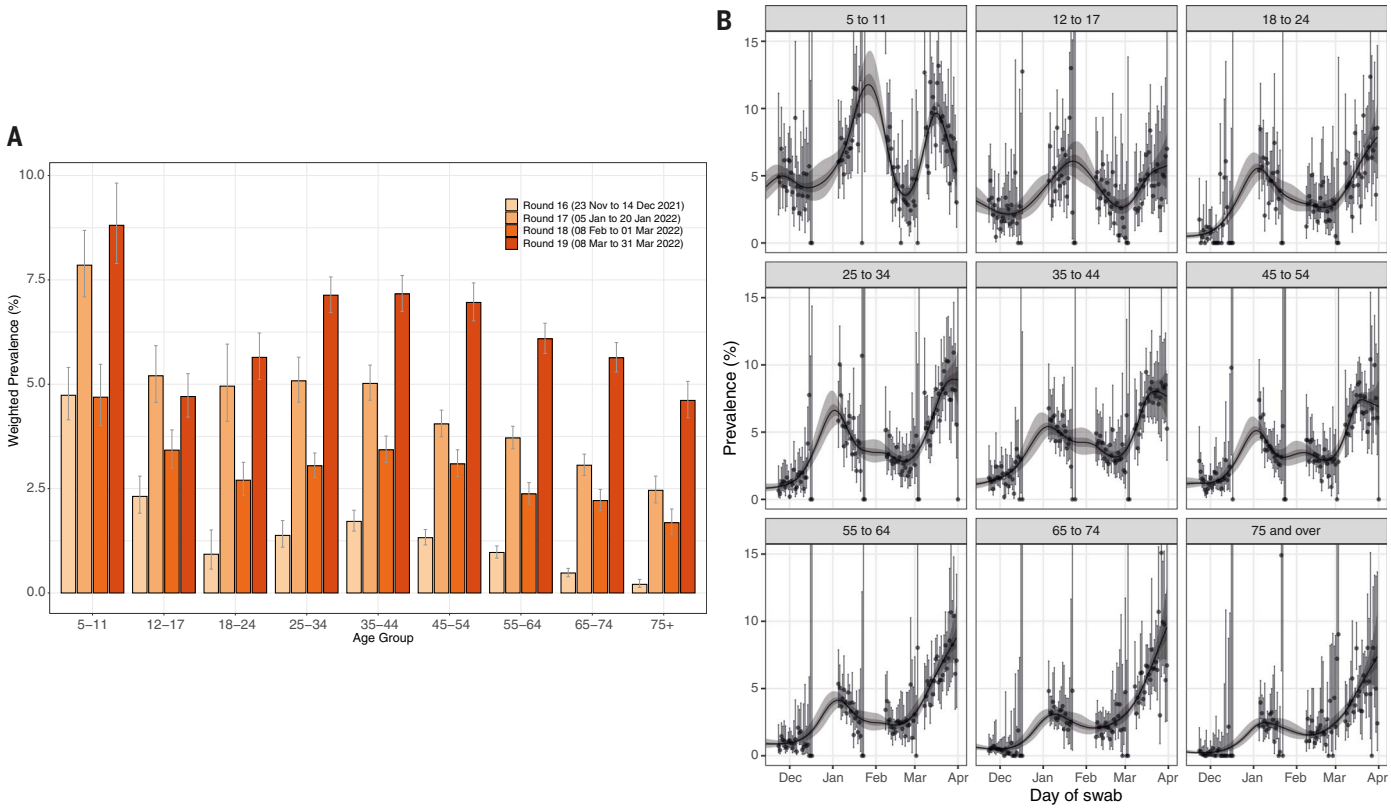


Fig. 2. Weighted prevalence by age. (**A**) Weighted prevalence of SARS-CoV-2 swab positivity by age group from round 16 (light orange) to round 19 (dark orange). Bars show the weighted prevalence point estimates, and the vertical lines represent the 95% CrIs. (**B**) Comparison of P-spline models fit to SARS-CoV-2 swab-positivity data from all rounds of REACT-1 for those aged 5 to 11, 12 to 17, 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and 75+ years. Shaded regions show 50% (dark shading) and 95% (light shading) posterior CrIs for the P-spline models. Results are presented for each day (*x* axis) of sampling for rounds 16 to 19, and the prevalence of swab positivity is shown (*y* axis). Weighted observations (dots) and 95% CrIs (vertical lines) are also shown.

to around 15.0%. Additionally, we used within-round random iterative method (rim) weighting (*32*) to correct the sample to be representative of the base population. During the 23 months of the study, we have adapted the way that samples were handled (for example, courier to post, no cold chain, inclusion of a multiplex PCR assay in the latter rounds). Although these changes may have introduced small effects into between-round comparisons, they should not have affected within-round trends.

We report an unprecedented and increasing prevalence of SARS-CoV-2 infections in England during March 2022. We observed Omicron twin peaks as BA.1 replaced Delta and BA.2 replaced BA.1 while, at the same time, society opened up, with all legal restrictions related to COVID-19 in England lifted as part of its "Living with COVID-19" strategy. These high rates of infections were associated with increasing hospitalizations and deaths due to COVID-19 in England during March 2022, but at much lower levels than in previous waves against a backdrop of high levels of vaccination in the population. These transmission dynamics in England may be relevant to the experience in the United States and other countries as BA.2 takes hold as the predominant variant worldwide.

Materials and methods

Study design

The REACT-1 study involved a series of cross-sectional surveys of random samples of the population of England at 5 years of age and older (*15*) carried out over 19 distinct rounds from 1 May 2020 to 31 March 2022. Those registering for the study were sent a self-administered throat and nose swab kit with instructions and asked to complete a questionnaire. In total, 2,512,797 participants had a valid test result for SARS-CoV-2 by RT-PCR across the 19 rounds of the study (Table 1) from among 14,036,117 individuals who were sent invitation letters, giving an overall response rate of 17.9% (ratio of completed tests to letters sent out). We focus here on the Omicron period spanning round 16 (23 November to 14 December 2021, N = 97,089), round 17 (5 to 20 January 2022, N = 102,174), round 18 (8 February to 1 March 2022, N = 94,950), and round 19 (8 to 31 March 2022, N = 109,181).

The sampling frame was the NHS general practitioner list of patients in England (covering almost the entire population), which includes name, address, age, and sex. Participants provided information on ethnicity, household size, occupation, symptoms, and other variables (*33*). We used residential post codes to link to an area-level index of multiple deprivation (an overall relative measure of deprivation) (*18*) and to urban or rural status (*17*). We used a multiplex that included influenza A and B for rounds 16 to 19; only the SARS-CoV-2 results are reported here.

Initially, we aimed to obtain approximately equal numbers of participants in each LTLA in England (N = 315), but from round 12 (20 May to 7 June 2021), we switched to obtaining a random sample in proportion to population size at the LTLA level. We used rim weighting (*32*) to provide prevalence estimates for the population of England as a whole, adjusting for age, sex, deciles of the index of multiple deprivation, LTLA counts, and ethnic group. Incentives were added to improve the response among underrepresented groups in rounds 18 and 19. For the return of a completed test, a gift voucher worth £10 was offered to those aged 13 to 17 and 35 to 44 years and £20 to those aged 18 to 34 years.

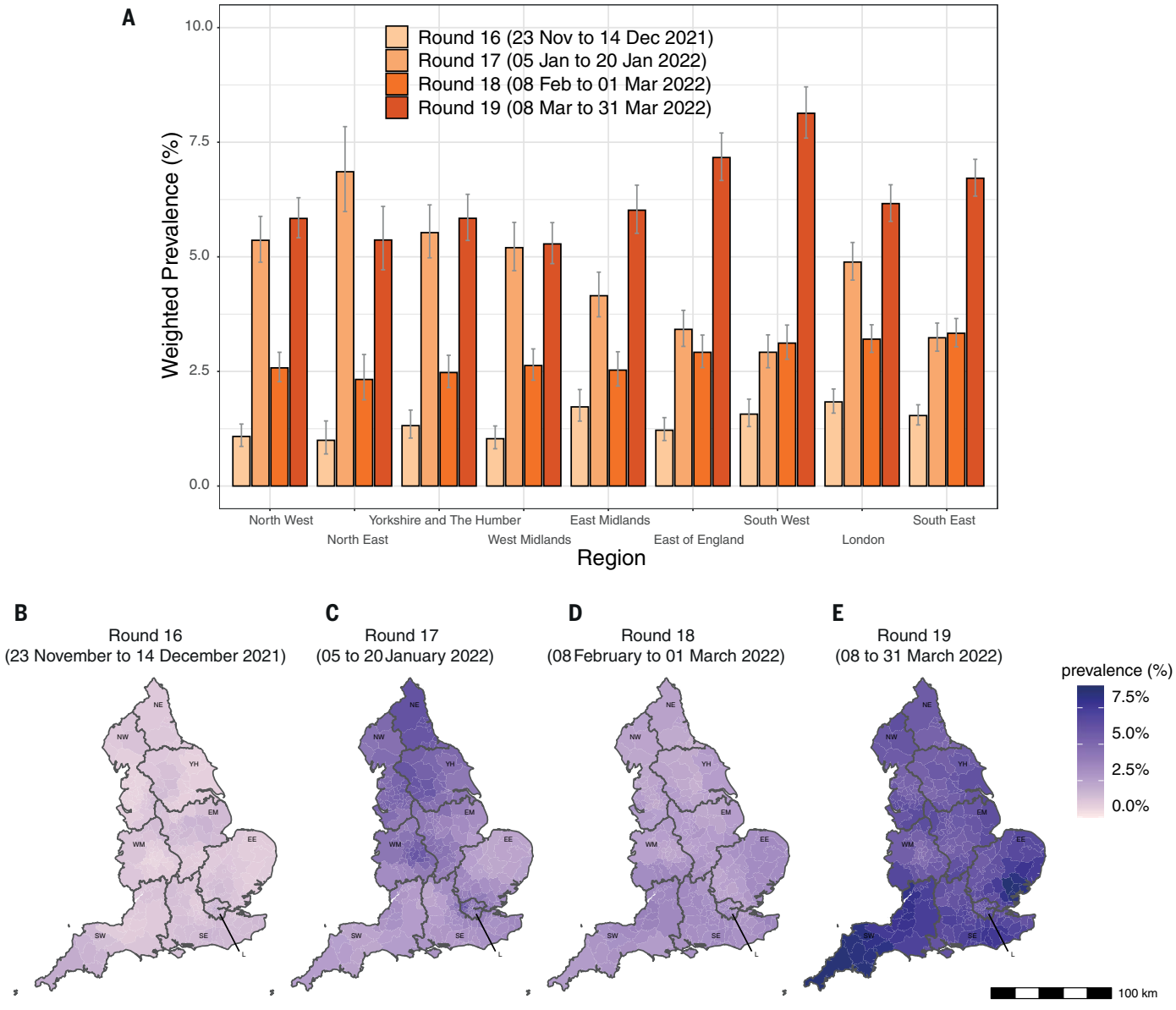


Fig. 3. SARS-CoV-2 swab positivity by region. (**A**) Weighted prevalence of SARS-CoV-2 swab positivity by region from round 16 (light orange) to round 19 (dark orange). Bars show the weighted prevalence point estimates, and the vertical lines represent the 95% CrIs. (**B** to **E**) Neighborhood smoothed-average SARS-CoV-2 swab-positivity prevalence by LTLA for rounds 16 (B), 17 (C), 18 (D), and 19 (E). Neighborhood prevalence was calculated from nearest neighbors (the median number of neighbors within 30 km in the study). Average neighborhood prevalence is displayed for each LTLA across the whole of England. Regions: EE, East of England; EM, East Midlands; L, London; NE, North East; NW, North West; SE, South East; SW, South West; WM, West Midlands; YH, Yorkshire and The Humber.

Table 3. Growth-rate advantage comparing the variant-specific daily growth rates for Alpha, Delta, Omicron, and other lineages. Point estimates and 95% CrIs (in parentheses) are reported along with the estimated time (in days) for the proportion of the lineage of interest to grow from 5 to 50% and from 5 to 95%. Results are presented for the transitions from wild type to Alpha, Alpha to Delta, Delta to Omicron, and non-BA.2 Omicron to BA.2 sublineages.

Lineages competing	Growth-rate advantage	Time for growth from 5 to 50% (days)	Time for growth from 5 to 95% (days)
Alpha versus wild type (November 2020 to April 2021)	0.029 (0.019, 0.042)	100.1 (157.8, 70.5)	200.2 (315.5, 141.0)
Delta versus Alpha (April to July 2021)	0.085 (0.070, 0.104)	34.8 (42.2, 28.4)	69.6 (84.4, 56.8)
Omicron versus Delta (December 2021 to January 2022)	0.207 (0.192, 0.224)	14.2 (15.4, 13.1)	28.5 (30.7, 26.3)
BA.2 versus non-BA.2 Omicron (January to April 2022)	0.104 (0.100, 0.109)	28.2 (29.4, 27.1)	56.4 (58.7, 54.2)

Up to round 13 (24 June to 12 July 2021), we collected dry swabs sent by courier to the laboratory on a cold chain, but from round 14 (9 to 27 September 2021, including 509 samples from 28 to 30 September), we switched to "wet" (saline) swabs, which, from round 14, were sent to the laboratory either by courier (no cold chain) or priority post and, from round 15 (October 19 to November 5 2021) onward, were sent by priority post only. Because of delays in the post for return of swabs, we include a small proportion of samples obtained after the nominated closing date for each of rounds 14 to 18.

A test result was positive if both N gene and E gene targets were detected or if N gene was detected with a cycle threshold (Ct) value less than 37.

Viral genome sequencing

We carried out viral genome sequencing (Quadram Institute, Norwich, UK) of positive samples with Ct ≤ 34 for either the E or N gene. We used the ARTIC protocol (*34*) (version 4 for rounds 16 and 17 and version 4.1 for rounds 18 and 19) for viral RNA amplification, CoronaHiT for preparation of sequencing libraries (*35*), and the ARTIC bioinformatics pipeline (*34*) and assigned lineages using Pangolin (v4.0 with pangolin-data v1.2.133) (*36*). RT-PCR was performed on 96 randomly chosen samples using the CDC assay (*37*) by the Quadram Institute as a secondary confirmation of the Ct values.

Data analyses

As noted, we used rim weighting (*32*) to estimate round-specific weighted prevalence and 95% CrIs. We used logistic regression to estimate the odds of testing positive by employment, ethnicity, household size, children in household, smoking and vaping status, urban or rural status, and deprivation, adjusting for age, region, and, subsequently, all other variables examined. We fit a Bayesian P-spline model (*38*, *39*) to the daily data to visualize temporal trends in swab positivity over the whole study period. Additionally, we fitted nine age group–specific P-splines (5 to 11, 12 to 17, 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and 75+ years), with the smoothing parameter obtained from the model fit to all the data. To fit the P-splines, we used a No-U-Turn Sampler in logit space (*40*), partitioning the data into ~5-day sections by regularly spaced knots and minimizing edge effects by adding further knots beyond the study period. Models were implemented and run using RStan (*41*). We used day of swabbing where reported; otherwise, we used day of pickup by courier or first post office scan where available. We guarded against overfitting by use of fourth-order basis splines (b-splines) over the knots, including a second-order random-walk prior distribution on the coefficients of the b-splines; the prior distribution penalized against changes in growth rate unless supported by the data (*39*).

We estimated r, the daily overall exponential growth or decay rate (for SARS-CoV-2 swab positivity), over the entire period of the study (beginning 1 May 2020) (*39*). The reproduction number, R, for the Omicron period was estimated assuming a gamma-distributed generation time with an Omicron-specific mean of 3.3 days and standard deviation of 3.5 days (shape n = 0.89 and rate β = 0.27) as (*42*)

$$R = \left(1 + \frac{r}{\beta}\right)^{n}$$

The use of prevalent swab-positivity data means that changes in the underlying exponential growth rate of new infections are not detected immediately. Instead, after a change in the underlying growth rate, the estimates move smoothly between the previous value and the more recent one, regardless of the underlying mixture of variants and sublineages.

We estimated the growth-rate advantage (a comparison of variant-specific growth rates) for the transition from wild type to Alpha, Alpha to Delta, Delta to Omicron, and Omicron BA.1 to BA.2 by fitting a Bayesian logistic regression model to the daily proportions of the competing variants. The daily relative growth rates in the log odds of Alpha, Delta, Omicron, and other lineages were estimated, assuming constant growth rates, using a Bayesian multinomial logistic regression model fit to the categorical outcome variable (Alpha, Delta, Omicron, other) over rounds 8 to 19, with Delta set as the reference category.

The time taken for the proportion of one lineage to increase from 5 to 50% was calculated assuming that only two lineages were present and using the pairwise difference in their growth rates, r^*, in the equation

$$T_{5\% \text{ to } 50\%} = \frac{\log\left(\frac{0.05}{1-0.05}\right) - \log\left(\frac{0.5}{1-0.5}\right)}{r^*}$$

For example, in calculating Delta's rise against Alpha, r^* would be the difference in growth rates of Alpha and Delta. Because of the assumed symmetry, the time of one lineage to increase from 5 to 95% is two times $T_{5\% \text{ to } 50\%}$.

To account for participants' proximity across LTLAs and the heterogeneous population density across LTLAs, we calculated smoothed prevalence per LTLA using a nearest neighbor approach. Briefly, for each LTLA, we estimated the median number of participants (M) within 30 km of each other. Then, for a random sample of 15 participants per LTLA, we calculated the prevalence of infection among the nearest M people. The smoothed prevalence by LTLA is then defined as the average of the 15 prevalence estimates in that area. Statistical analyses were performed with R software, version 4.0.5.

Mobility data

Daily data on mobility (transit, driving, and walking) were downloaded for England from Apple Mobility Trends Report (*43*). Seven-day moving averages, relative to the maximum 7-day average between 1 May 2020 and 31 March 2022, were plotted on the fourth of the 7 days. Days were defined from midnight to midnight, US Pacific time. Random rotating identifiers, rather than Apple IDs, are used for data sent from Apple users to the Apple Maps service. Thus, no profiles are collected on individual movements. Because Apple Maps thus has no demographic data on users, it is not possible to assess the representativeness of the mobility data provided.

Lockdown dates and restrictions in England

In Fig. 1D, gray-shaded regions represent periods when lockdown was implemented in England. The following dates were used as start and end dates (from 1 May 2020 to 31 March 2022):

23 March 2020: The first national lockdown was announced (*44*).

23 June 2020: The Prime Minister announced key changes to lockdown restrictions (*45*).

5 November 2020: The second national lockdown in England was announced (*46*).

2 December 2020: The second national lockdown in England ended (*47*) after 4 weeks. England moved to a stricter three-tiered system of restrictions.

6 January 2021: The third national lockdown in England was announced (*48*).

8 March 2021: England started a phased release of lockdown regulations (*49*).

18 July 2021: Lockdown laws ceased to be in force (*50*).

24 February 2022: All domestic legal restrictions concerning COVID-19 in England were removed (*29*).

1 April 2022: All remaining restrictions in England were removed (*29*).

The easing of restrictions during or after lockdown was not complete at a single time point, so these dates up to 1 April 2022 should not be regarded as representing presence or absence of all restrictions.

REFERENCES AND NOTES

1. P. Elliott *et al.*, Rapid increase in Omicron infections in England during December 2021: REACT-1 study. *Science* **375**, 1406–1411 (2022). doi: 10.1126/science.abn8347; pmid: 35133177
2. UK Health Security Agency, "SARS-CoV-2 variants of concern and variants under investigation in England: Technical briefing 33" (UK Health Security Agency, 2021); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1043680/technical-briefing-33.pdf.
3. P. Elliott *et al.*, Post-peak dynamics of a national Omicron SARS-CoV-2 epidemic during January 2022. medRxiv

2022.02.03.22270365 [Preprint] (2022); doi: 10.1101/2022.02.03.22270365
4. M. Chadeau-Hyam *et al*., The Omicron SARS-CoV-2 epidemic in England during February 2022. medRxiv 2022.03.10.22272177 [Preprint] (2022); .doi: 10.1101/2022.03.10.22272177
5. O. Eales *et al*., The new normal? Dynamics and scale of the SARS-CoV-2 variant Omicron epidemic in England. medRxiv 2022.03.29.22273042 [Preprint] (2022); doi: 10.1101/2022.03.29.22273042
6. World Health Organization (WHO), WHO coronavirus (COVID-19) dashboard; https://covid19.who.int/.
7. US Centers for Disease Control and Prevention (CDC), COVID data tracker; https://covid.cdc.gov/covid-data-tracker/.
8. T. Nyberg *et al*., Comparative analysis of the risks of hospitalization and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: A cohort study. *Lancet* **399**, 1303–1312 (2022). doi: 10.1016/S0140-6736(22)00462-7; pmid: 35305296
9. M. Hutton, "Hong Kong Covid crisis: why is the death rate so high?" *Guardian*, 18 March 2022; https://www.theguardian.com/world/2022/mar/18/hong-kong-covid-crisis-why-is-the-death-rate-so-high.
10. V. C.-C. Cheng *et al*., Rapid spread of SARS-CoV-2 Omicron subvariant BA.2 in a single-source community outbreak. *Clin. Infect. Dis.*Ciac203 (2022). doi: 10.1093/cid/ciac203; pmid: 35271728
11. O. Dyer, Covid-19: Lockdowns spread in China as omicron tests "zero covid" strategy. *BMJ* **376**, o859 (2022). doi: 10.1136/bmj.o859; pmid: 35361680
12. Centre for Health Protection, Department of Health, The Government of Hong Kong Special Administrative Region, Latest situation of COVID-19 (as of 14 May 2022); https://www.chp.gov.hk/files/pdf/local_situation_covid19_en.pdf [accessed 14 May 2022].
13. Centre for Health Protection, Department of Health, The Government of Hong Kong Special Administrative Region, Statistics on 5th wave of COVID-19; https://www.covidvaccine.gov.hk/pdf/5th_wave_statistics.pdf.
14. GOV.UK, UK summary: The official UK government website for data and insights on coronavirus (COVID-19); https://coronavirus.data.gov.uk/.
15. S. Riley *et al*., Real-time Assessment of Community Transmission (REACT) of SARS-CoV-2 virus: Study protocol. *Wellcome Open Res.* **5**, 200 (2021). doi: 10.12688/wellcomeopenres.16228.1; pmid: 33997297
16. A. J. Kucharski, C. Cohen, Effective surveillance of variants. *Science* **375**, 1349–1350 (2022). doi: 10.1126/science.abo4257; pmid: 35324314
17. Office for National Statistics, 2011 rural/urban classification; https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2011ruralurbanclassification.
18. Office for National Statistics, Mapping income deprivation at a local authority level: 2019; https://www.ons.gov.uk/releases/mappingincomedeprivationatalocalauthoritylevel2019.
19. P. Elliott *et al*., Exponential growth, high prevalence of SARS-CoV-2, and vaccine effectiveness associated with the Delta variant. *Science* **374**, eabl9551 (2021). doi: 10.1126/science.abl9551; pmid: 34726481
20. Office for National Statistics, Coronavirus (COVID-19) latest insights: Infections; https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19latestinsights/infections.
21. R. Yapp, M. Bracher, Coronavirus (COVID-19) infection survey, UK: 14 April 2022; https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/14april2022.
22. S. Riley *et al*., Resurgence of SARS-CoV-2: Detection by community viral surveillance. *Science* **372**, 990–995 (2021). doi: 10.1126/science.abf0874; pmid: 33893241
23. UK Health Security Agency, "SARS-CoV-2 variants of concern and variants under investigation in England: Technical briefing 34" (UK Health Security Agency, 2022); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1048395/technical-briefing-34-14-january-2022.pdf.
24. F. P. Lyngse *et al*., Transmission of SARS-CoV-2 Omicron VOC subvariants BA.1 and BA.2: Evidence from Danish households. medRxiv 2022.01.28.22270044 [Preprint] (2022); doi: 10.1101/2022.01.28.22270044
25. R. P. Bhattacharyya, W. P. Hanage, Challenges in inferring intrinsic severity of the SARS-CoV-2 omicron variant. *N. Engl. J. Med.* **386**, e14 (2022). doi: 10.1056/NEJMp2119682; pmid: 35108465
26. UK Health Security Agency, "SARS-CoV-2 variants of concern and variants under investigation in England: Technical briefing 39" (UK Health Security Agency, 2022); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1063424/Tech-Briefing-39-25March2022_FINAL.pdf.
27. M. van Algemene Zaken, Slowing the spread of the Omicron variant: lockdown in the Netherlands (2021); https://www.government.nl/ministries/ministry-of-health-welfare-and-sport/news/2021/12/18/slowing-the-spread-of-the-omicron-variant-lockdown-in-the-netherlands.
28. E. Anthes, A. Ghorayshi, "In omicron hot spots, hospitals fill up, but I.C.U.s may not," *New York Times*, 4 January 2022; https://www.nytimes.com/2022/01/04/health/covid-omicron-hospitalizations.html.
29. GOV.UK, Prime Minister's Office, Prime Minister sets out plan for living with COVID (2022); https://www.gov.uk/government/news/prime-minister-sets-out-plan-for-living-with-covid.
30. GOV.UK, Cabinet Office, COVID-19 response: Living with COVID-19 (2022); https://www.gov.uk/government/publications/covid-19-response-living-with-covid-19.
31. National Health Service (NHS), "COVID-19 vaccination statistics" (NHS, 2022); https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2022/03/COVID-19-weekly-announced-vaccinations-31-March-2022.pdf.
32. T. Sharot, Weighting survey results. *J. Market Res. Soc.* **28**, 269–284 (1986).doi: 10.1186/s13073-021-00839-5; pmid: 33563320
33. Imperial College London, REACT 1 study materials; https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/for-researchers/react-1-study-materials/.
34. connor-lab/ncov2019-artic-nf. Github (2020); https://github.com/connor-lab/ncov2019-artic-nf.
35. D. J. Baker *et al*., CoronaHiT: High-throughput sequencing of SARS-CoV-2 genomes. *Genome Med.* **13**, 21 (2021). doi: 10.1186/s13073-021-00839-5; pmid: 33563320
36. cov-lineages/pangolin. Github (2022); https://github.com/cov-lineages/pangolin.
37. US Centers for Disease Control and Prevention (CDC), Research use only 2019-novel coronavirus (2019-nCoV) real-time RT-PCR primers and probes; https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html.
38. S. Lang, A. Brezger, Bayesian P-splines. *J. Comput. Graph. Stat.* **13**, 183–212 (2004). doi: 10.1198/1061860043010
39. O. Eales *et al*., Appropriately smoothing prevalence data to inform estimates of growth rate and reproduction number. medRxiv 2022.02.04.22270426 [Preprint] (2022); doi: 10.1101/2022.02.04.22270426
40. M. D. Hoffman, A. Gelman, The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).
41. Stan Development Team, Rstan: The R interface to Stan. R package version 2.18.2 (2018); http://mc-stan.org.
42. S. Abbott, K. Sherratt, M. Gerstung, S. Funk, Estimation of the test to test distribution as a proxy for generation interval distribution for the Omicron variant in England. medRxiv 2022.01.08.2226892 [Preprint] (2022); doi: 10.1101/2022.01.08.22268920
43. Apple Maps, COVID-19 – Mobility trends reports; https://covid19.apple.com/mobility [accessed 1 April 2022].
44. GOV.UK, Prime Minister's Office, Prime Minister's statement on coronavirus (COVID-19): 23 March 2020 (2020); https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-23-march-2020.
45. GOV.UK, Prime Minister's Office, PM announces easing of lockdown restrictions: 23 June 2020 (2020); https://www.gov.uk/government/news/pm-announces-easing-of-lockdown-restrictions-23-june-2020.
46. GOV.UK, Prime Minister's Office, Prime Minister's statement on coronavirus (COVID-19): 5 November 2020 (2020); https://www.gov.uk/government/speeches/prime-ministers-statement-on-coronavirus-covid-19-5-november-2020.
47. GOV.UK, Prime Minister's Office, Prime Minister's statement on coronavirus (COVID-19): 2 December 2020 (2020); https://www.gov.uk/government/speeches/prime-ministers-statement-on-coronavirus-covid-19-2-december-2020.
48. GOV.UK, Prime Minister's Office, Prime Minister's statement to the House of Commons on COVID-19 regulations: 6 January 2021 (2021); https://www.gov.uk/government/speeches/prime-ministers-statement-to-the-house-of-commons-on-covid-19-regulations-6-january-2021.
49. GOV.UK, Cabinet Office, COVID-19 Response – Spring 2021 (Summary): Roadmap out of lockdown (2021); https://www.gov.uk/government/publications/covid-19-response-spring-2021/covid-19-response-spring-2021-summary.
50. UK Parliament, Coronavirus: A history of English lockdown laws (2021); https://commonslibrary.parliament.uk/research-briefings/cbp-9068/.
51. O. Eales, M. Whitaker, H. Wang, c97sr, C. Walters, R. FitzJohn, K. Ainslie, mathzero/reactidd: Elliott et al, "Twin peaks...." paper, Science 2022. Zenodo (2022); doi: 10.5281/zenodo.6550327
52. GISAID; https://www.gisaid.org/.

ACKNOWLEDGMENTS

We thank key collaborators on this work: K. Beaver, S. Clemens, G. Welch, N. Gilby, K. Ward, G. Pantelidou, and K. Pickering at Ipsos; G. Fontana and J. Alford at the Institute of Global Health Innovation at Imperial College London; E. Johnson, R. Elliott, and G. Blakoe at the School of Public Health, Imperial College London; and N.-F. Alikhan at the Quadram Institute, Norwich, UK. We also thank North West London Pathology and Public Health England (now UKHSA) for help in calibration of the laboratory analyses, Patient Experience Research Centre at Imperial College London and the REACT Public Advisory Panel, NHS Digital for access to the NHS register, and the Department of Health and Social Care for logistical support. **Funding:** This study was funded by the Department of Health and Social Care in England. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of this manuscript. P.E. is director of the Medical Research Council (MRC) Centre for Environment and Health (MR/L01341X/1 and MR/S019669/1). P.E. acknowledges support from Health Data Research UK (HDR UK); the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre; NIHR Health Protection Research Units in Chemical and Radiation Threats and Hazards, and Environmental Exposures and Health; the British Heart Foundation Centre for Research Excellence at Imperial College London (RE/18/4/34215); and the UK Dementia Research Institute at Imperial College London (MC_PC_17114). N.S. acknowledges support from the Oxford-Radcliffe Scholarship from University College, Oxford, and the Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training (CDT) in Modern Statistics and Statistical Machine Learning (Imperial College London and University of Oxford). We thank A. Maslov for supporting this research through studentship support for N.S. A.J.P. acknowledges support from the Biotechnology and Biological Sciences Research Council (BB/R012504/1). H.War. acknowledges support from an NIHR Senior Investigator Award, the Wellcome Trust (205456/Z/16/Z), and the NIHR Applied Research Collaboration (ARC) North West London. J.E. is an NIHR academic clinical fellow in infectious diseases. G.S.C. is supported by an NIHR professorship. C.A.D. acknowledges support from the MRC Centre for Global Infectious Disease Analysis, the NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, and the NIHR-funded Vaccine Efficacy Evaluation for Priority Emerging Diseases (PR-OD-1017-20007). M.C.-H. and B.B. acknowledge support from Cancer Research UK, Population Research Committee Project grant "Mechanomics" (grant no. 22184 to M.C.-H.). M.C.-H. acknowledges support from the H2020-EXPANSE (Horizon 2020 grant no. 874627) and H2020-LongITools (Horizon 2020 grant no. 874739). **Ethics:** We obtained research ethics approval from the South Central-Berkshire B Research Ethics Committee (IRAS ID: 283787). Notification of favorable opinion and brief summary of the protocol are available at https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/react1-covid-19-uph/. **Public involvement:** A public advisory panel provides input into the design, conduct, and dissemination of the REACT research program. **Author contributions:** P.E., C.A.D., and M.C.-H. are corresponding authors. Conceptualization: P.E., C.A., P.J.D., D.A., W.B., H.War., A.D., G.S.C., M.C.-H., C.A.D.; Methodology: P.E., O.E., N.S., D.T., B.B., H.Wan., P.J.D., A.P., H.War., C.A.D., M.C.-H.; Software: O.E., N.S., D.T., B.B., H.Wan., M.C.-H.; Validation: O.E., H.Wan., G.T.; Formal analysis: O.E., N.S., D.T., B.B., H.Wan., M.W., A.J.P., A.J.T., M.C.-H.; Investigation: P.E., A.J.T., M.C.-H.; Resources: P.E., A.J.P.; Data curation: O.E., N.S., D.T., B.B., M.W., A.J.P., A.J.T.; Writing – original draft: P.E., M.C.-H., C.A.D.; Writing – review & editing: P.E., O.E., N.S., J.E., C.A., P.J.D., A.J.P., A.J.T., D.A., W.B., G.T., H.War., A.D., G.S.C., C.A.D.; Visualization: O.E., N.S., D.T., B.B., H.Wan., M.W., M.C.-H.; Supervision: P.E., C.A., D.A., W.B., H.War., A.D., G.S.C., M.C.-H., C.A.D.; Project administration: P.E., C.A., W.B., H.War., G.S.C., C.A.D., M.C.-H.; Funding acquisition: P.E., C.A., D.A., W.B., H.War.,

A.D., G.S.C., C.A.D. All authors revised the manuscript for important intellectual content and approved the submission of the manuscript. P.E., C.A.D., and M.C.-H. had full access to the data and take responsibility for the integrity of the data and the accuracy of the data analysis and for the decision to submit for publication. **Competing interests**: A.D. is chairman of the Health Security and Pre-Emptive Medicine Initiative, Flagship Pioneering UK plc, and has no conflict of interest to declare. M.C.-H. holds shares in the O-SMOSE company. Consulting activities conducted by the company are independent of the present work, and M.C.-H. has no conflict of interest to declare. All other authors have no competing interests to declare. **Data and materials availability**: Access to REACT-1 individual-level data is restricted to protect participants' anonymity. Summary statistics, descriptive tables, and code from the current REACT-1 study are available at https://github.com/mrc-ide/reactidd (*51*). REACT-1 study materials are available for each round at https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/for-researchers/react-1-study-materials/. Sequence read data are available without restriction from the European Nucleotide Archive at https://www.ebi.ac.uk/ena/browser/view/PRJEB37886, and consensus genome sequences are available from the Global Initiative on Sharing All Influenza Data (GISAID) (*52*). **License information:** This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq4411
Supplementary Text
Tables S1 to S3
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 8 April 2022; accepted 20 May 2022
Published online 24 May 2022
10.1126/science.abq4411

RESEARCH ARTICLE

ORGANIC CHEMISTRY

Nickel-catalyzed hydrogenative coupling of nitriles and amines for general amine synthesis

Vishwas G. Chandrashekhar, Wolfgang Baumann, Matthias Beller[1]*, Rajenahally V. Jagadeesh[1]*

Efficient and general methods for the synthesis of amines remain in high demand in the chemical industry. Among the many known processes, catalytic hydrogenation is a cost-effective and industrially proven reaction and currently used to produce a wide array of such compounds. We report a homogeneous nickel catalyst for hydrogenative cross coupling of a range of aromatic, heteroaromatic, and aliphatic nitriles with primary and secondary amines or ammonia. This general hydrogenation protocol is showcased by straightforward and highly selective synthesis of >230 functionalized and structurally diverse amines including pharmaceutically relevant and chiral products, as well as ^{15}N-isotope labeling applications.

Over the past century, commercial production of amines has continuously grown and now exceeds 6 million metric tons per year of valuable fine and bulk chemicals (*1*–*5*). Additionally, over the next decade the amine market is predicted to increase by ~8% annually worldwide as a result of growing global demand (*4*, *5*). Today amines are widely used in the manufacture of consumer products, advanced chemicals, pharmaceuticals, agrochemicals, and polymers (*1*–*5*). For example, the majority of agrochemicals and >75% of the top 200 selling drugs contain amine moieties, which often play crucial roles in their modes of action (*3*–*5*).

Because of the demand for amines there is a constant need to improve their synthesis by applying more economical and sustainable methodologies. Currently, catalytic hydrogenation reactions of nitriles (*6*–*10*) and nitroarenes (*11*, *12*) are mainly used in the industry for making primary aliphatic and aromatic amines whereas reductive amination of carbonyl compounds (*13*–*16*), amination of alcohols (*17*, *18*), substitution of halogenated compounds with amines (*19*, *20*), and hydroamination of olefins (*21*, *22*) are mostly used for making secondary and tertiary amines. Among all the known methodologies, catalytic hydrogenation of nitriles represents a cost-effective and atom-economical industrial process to produce primary benzylic and aliphatic amines (*6*–*10*). This methodology plays a key role in bulk industrial production of nylon (through 1,6-hexamethylenediamine), high-performance plastics (1,4-diaminobutane), surfactants (alkyl dimethyl amines), emulsifiers (e.g., aminoethylpiperazine), and bioactive molecules. Many aromatic, heterocyclic, and aliphatic nitriles are easily accessible and produced on a commercial scale. Additionally, more than 100 naturally occurring nitriles have been isolated from terrestrial and marine sources (*23*, *24*). Hence the direct catalytic hydrogenation of nitriles to produce primary amines continues to attract attention for both academic and industrial synthesis (*6*–*10*). By contrast, the synthesis of diverse secondary and tertiary amines—including *N*-methylamines from nitriles—by cross coupling is highly challenging and does not typically proceed with high selectivity (*25*–*30*). As shown in Fig. 1, the hydrogenation of nitrile **1** generates primary imine **2**. This intermediate is reduced to form the primary amine **3**, which can further react with **2** to produce the secondary imine **4**. Subsequent hydrogenation leads to secondary amine **5** (homocoupling pathway). Following a similar reaction pathway, the corresponding tertiary amine **6** can be formed. Alternatively, in the presence of different primary or secondary amines **7** and **8**, imine **2** can form imines **9** and **10**, which produce secondary or tertiary amines **11** and **12** (cross-coupling pathway). Further, secondary amine **11** can also react with **2** to produce **13**. Consequently, five different amines can result from the coupling reaction of a given nitrile with a primary or secondary amine. In addition, the conversion of aromatic nitriles to give various heterocycles, especially 2,4,5-substituted imidazoles **14** under hydrogenation conditions, has been described (*6*).

To perform reductive cross coupling of nitriles with amines in an efficient and selective manner the design of suitable catalyst systems is crucial. To obtain desired products **11** or **12**, the rate for the hydrogenation step of the initially formed primary imine **2** should be slow compared with the corresponding secondary imines. In addition, reversible nucleophilic addition of the amine to the imine and elimination of ammonia should be fast. Ideally, the applied catalyst should be based on an abundant, relatively inexpensive 3d-metal. We present a solution for these challenges as well as a general homogeneous nickel catalyst system for the efficient reductive cross coupling of nitriles with ammonia or amines, which allows for the synthesis of functionalized and structurally diverse primary, secondary, and tertiary benzylic, heterocyclic, and aliphatic amines including *N*-methylamines, along with more complex drug targets.

Design of catalyst and reaction optimization

In recent years multidentate phosphines have become privileged ligand scaffolds for the preparation of defined metal complexes with specific catalytic activity for demanding hydrogenation reactions (*31*–*42*). In this respect, the characteristic behavior of such catalysts for the hydrogenation of esters to ethers (*34*, *35*), carbon dioxide to methanol (*36*–*38*), and carboxylic acids to alcohols (*39*, *40*) as well as amides to amines (*41*, *42*) has been demonstrated. On the basis of these works, we investigated the performance of 3d-metals in combination with selected di-, tri-, and tetradentate phosphines for cross coupling of 4-methylbenzonitrile **15** with 4-anisidine **16** to produce 4-methoxy-*N*-(4-methylbenzyl)aniline **17** in the presence of molecular hydrogen as the model reaction (Fig. 2 and table S1). The coupling of **16** is also demanding because of the lower nucleophilicity compared with the benzylic amine **18**, which leads to self-coupling. Initially, we tested Fe-, Mn-, Co-, and Ni-salts in combination with bis-(2-diphenylphosphinoethyl)phenylphosphine (linear-triphos **L1**), tris[2-(diphenylphosphino) ethyl]phosphine (tetraphos **L2**), and 4,5-bis (diphenylphosphino)-9,9-dimethylxanthene (xantphos **L3**). To increase the reactivity of the resulting complexes, metal salts with non-coordinating anions, e.g., tetrafluoroborates and triflates, were tested. In general, compared with Co and Ni the in situ generated Fe and Mn complexes showed very low or no catalytic activity in the model reaction (Fig. 2 and table S1, entries 1 to 9). Comparing nickle and cobalt complexes, excellent activity and selectivity for the formation of **17** (91% yield) was observed in the presence of Ni-**L1** (Fig. 2 and table S1, entries 16 and 17). With the exception of $Ni(OTf)_2$ and $Ni(BF_4)_2{\cdot}6H_2O$, other nickel salts in the presence of **L1** gave significantly less active catalysts (table S1, entries 28 to 29). However, in the latter cases the high activity was restored in the presence of $Zn(OTf)_2$ (table S1, entry 30). As a result of the superior behavior of nickel, other ligands (**L4** to **L9**) were also tested with $Ni(OTf)_2$ in the model reaction. Notably, the in situ–generated Ni-complexes with **L4**, **L5**, and **L8** gave moderate to good yields (40 to 62% of **17**), whereas **L7** and **L9** exhibited only poor selectivity (16 to

[1]Leibniz-Institut für Katalyse e.V., Albert-Einstein-Str. 29a, D-18059 Rostock, Germany.
*Corresponding author. Email: matthias.beller@catalysis.de (M.B.); jagadeesh.rajenahally@catalysis.de (R.V.J.)

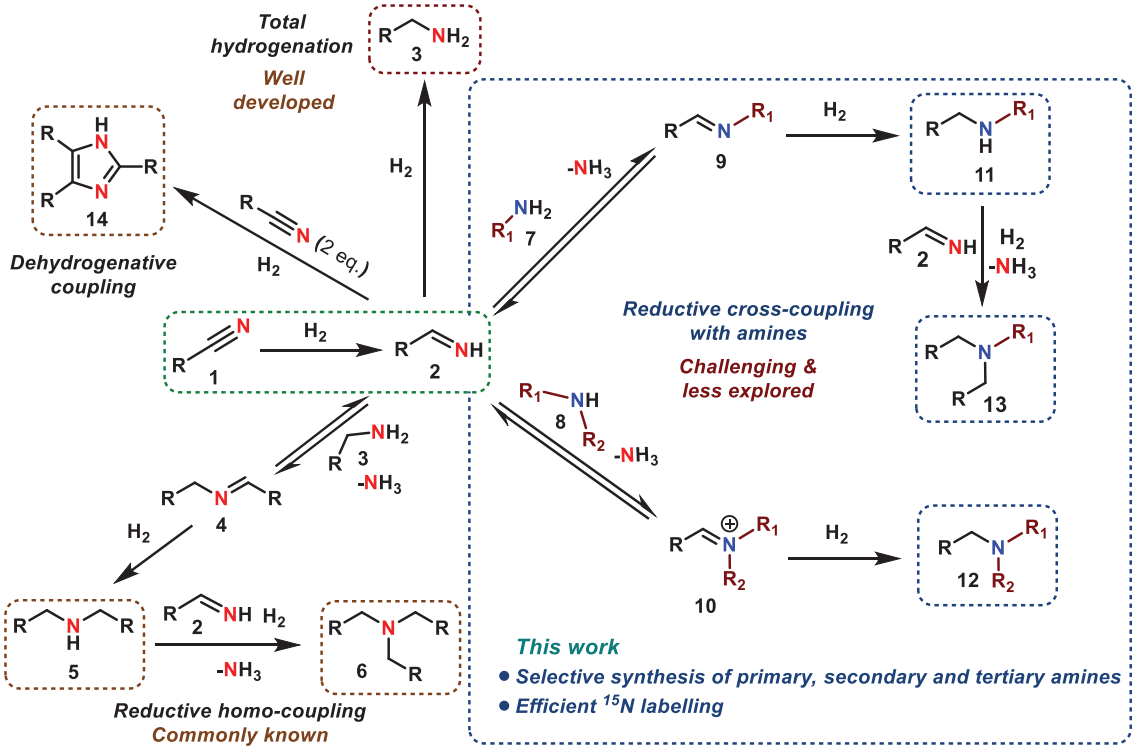


Fig. 1. Catalytic reductive cross coupling of nitriles with primary or secondary amines or ammonia and formation of different products.

24% of **17**), and **L6** provided no desired product (Fig. 2 and table S1, entries 22 to 27).

After optimization of key parameters (e.g., catalyst loading, temperature, H_2 pressure, solvent) applying Ni-**L1** produced an excellent yield (91% of **17**) in trifluoroethanol (TFE) with 40 bar of hydrogen at 100°C (tables S2 and S3). Though most of the catalyst evaluation studies were performed with in situ–generated catalysts we also prepared the defined $Ni(OTf)_2$-**L1** complex, which has been characterized by single-crystal x-ray diffraction, nuclear magnetic resonance (NMR), high-resolution mass spectrometry, and infrared spectroscopy. X-ray analysis of complex **A** revealed the coordination of one phosphine ligand (1:1 ratio of Ni: **L1**) and one triflate group to the Ni center (supplementary materials, section S2). This isolated complex (complex **A**) exhibited similar activity and selectivity to that of the in situ system (table S2, entry 4).

Monitoring the progress of the model reaction at different time intervals in the presence of the optimal Ni-**L1** system revealed initial formation of the secondary imine **19** which was subsequently reduced in a highly selective manner to the desired amine **17**. As shown in fig. S1 the reaction is completed after 12 hours and the hydrogenation of **19** is the rate-determining step. Looking at the likely elementary reaction steps (fig. S2) we hypothesize that the catalytic reduction of nitrile **15** takes place first to generate primary imine **15A**. This unstable intermediate can immediately couple with *p*-anisidine **16** to form aminal **15B**, which yields the stable secondary imine **19** upon elimination of ammonia. Then, **19** undergoes reduction to give the desired secondary amine **17**. The first hydrogenation step, as well as the imine formation, is fast compared with subsequent formation of **17**. The high chemoselectivity can be explained by the slow hydrogenation of the primary imine **15A** relative to the more rapid cross-coupling step. Thus, other unwanted side products such as the symmetrical secondary imine **20** and amine **21** are minimized.

Hydrogenative cross coupling of nitriles with primary and secondary amines

After identifying the optimal catalyst system Ni-**L1** in the benchmark reaction, we explored its applications for the synthesis of different classes of amines (Figs. 3 to 6). First, we performed hydrogenative cross-coupling reactions of substituted and functionalized aromatic, heterocyclic, and aliphatic nitriles with *p*-anisidine or aniline to produce the corresponding amines in good to excellent selectivity and yields (Fig. 3A). For example, sterically hindered 2,6-dimethylbenzonitrile and various halogenated benzonitriles reacted smoothly in up to 90% yield (Fig. 3A; products **24** and **26** to **33**). Similarly, di- and tri-substituted benzonitriles gave the respective secondary amines in up to 90% yield (Fig. 3A; products **30** to **33** and **36** to **37**). More challenging substrates such as benzonitriles containing reducible functional groups reacted smoothly to yield the corresponding amines (Fig. 3A; products **39** to **42**). Specifically, functional groups such as (thio)ether, ester, amide, and boronic acid ester were compatible. Next, heterocyclic amines based on pyridine, furan, benzodioxole, and 1,3-dihydroisobenzofuran were prepared in 80 to 92% yield (Fig. 3A; products **43** to **48**). Further, various phenylalkyl and aliphatic nitriles efficiently underwent cross coupling with *p*-anisidine (Fig. 3A; products **49** to **55**). A wide variety of structurally diverse primary amines also proved compatible (Fig. 4B), with 4-toluonitrile as the main coupling partner (Fig. 3B; products **56** to **95**). Trans-4-aminocyclohexanol, morpholine, and piperazine, which represent important motifs in drugs and natural products, were easily coupled and gave the corresponding amines in high yields (Fig. 3B; products **88, 91**, and **92**). Beyond primary amines, various secondary amines including indole were coupled with 4-tolunitrile and produced corresponding tertiary amines in up to 94% yields (Fig. 3B; products **89** to **95**). In addition, α-methyl amines also reacted smoothly with different nitriles to provide the desired products in up to 91% yield (Fig. 3C; products **96** to **100**). Apart from all these successful examples, we specifically looked for substrates that showed lower or no reactivity. These compounds are listed in figs. S3 and S4. As shown

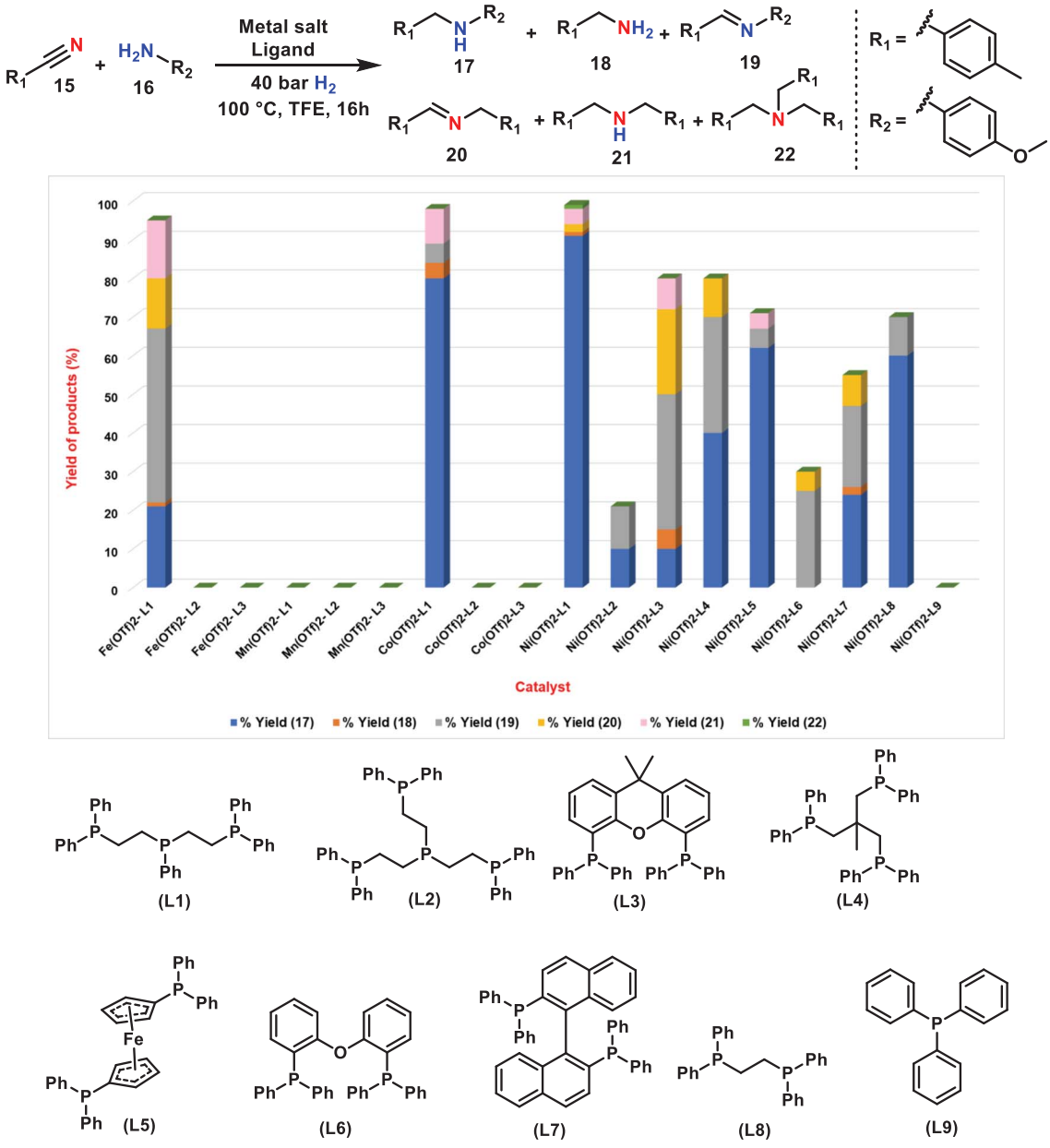


Fig. 2. Hydrogenative coupling of 4-methylbenzonitrile and 4-anisidine using in situ-generated metal-phosphine complexes. Reaction conditions: 0.5 mmol *p*-tolunitrile, 0.75 mmol *p*-anisidine, 4 mol % metal salt, 5 mol % ligand, 40 bar H_2, 3 mL TFE, 100°C, 16 hours, yields determined by GC with 4-flourotoluene as standard.

in these figures, less nucleophilic aromatic secondary amines, as well as sterically hindered nitriles and amines, did not react well and gave poor yields of the desired products. Furthermore, some substrates with functional groups including iodo, nitro, aldehyde, and alkyne substrates were not well tolerated.

Simple methyl- and dimethylamine constitute inexpensive bulk chemicals which are produced on 4 million metric tons per annum (*43*); however, their widespread use in organic synthesis is hampered by their physical properties (gaseous at room temperature) and strong binding affinity toward metal complexes. Nevertheless, the Ni-**L1** system exhibited good to excellent activity and selectivity for the reaction of aromatic nitriles with dimethyl- and monomethylamine and provided the corresponding *N*-methylated products in up to 92% yields (Fig. 4A; products **101** to **139**). This class of compounds is of particular importance as the *N*-methyl group represents a ubiquitous motif in many drugs and biomolecules. Notably, many aliphatic nitriles, including dinitriles, also underwent selective cross coupling with methyl- and dimethylamine (Fig. 4A; products **120** to **124** and **137** to **139**).

Hydrogenative cross coupling of nitriles with ammonia

After the successful synthesis of a plethora of secondary and tertiary amines by hydrogenative cross coupling we turned our interest to ammonia as an amine source. Products in this case are primary amines, which are typically prepared by traditional hydrogenation of nitriles. Simple hydrogenation of *p*-toluonitrile **15** under standard conditions in the presence of Ni-**L1** did not provide the desired 4-methylbenzylamine **18**. Instead, the formation of the self-coupled secondary amine **21** as the main product in 85% yield and 10 to

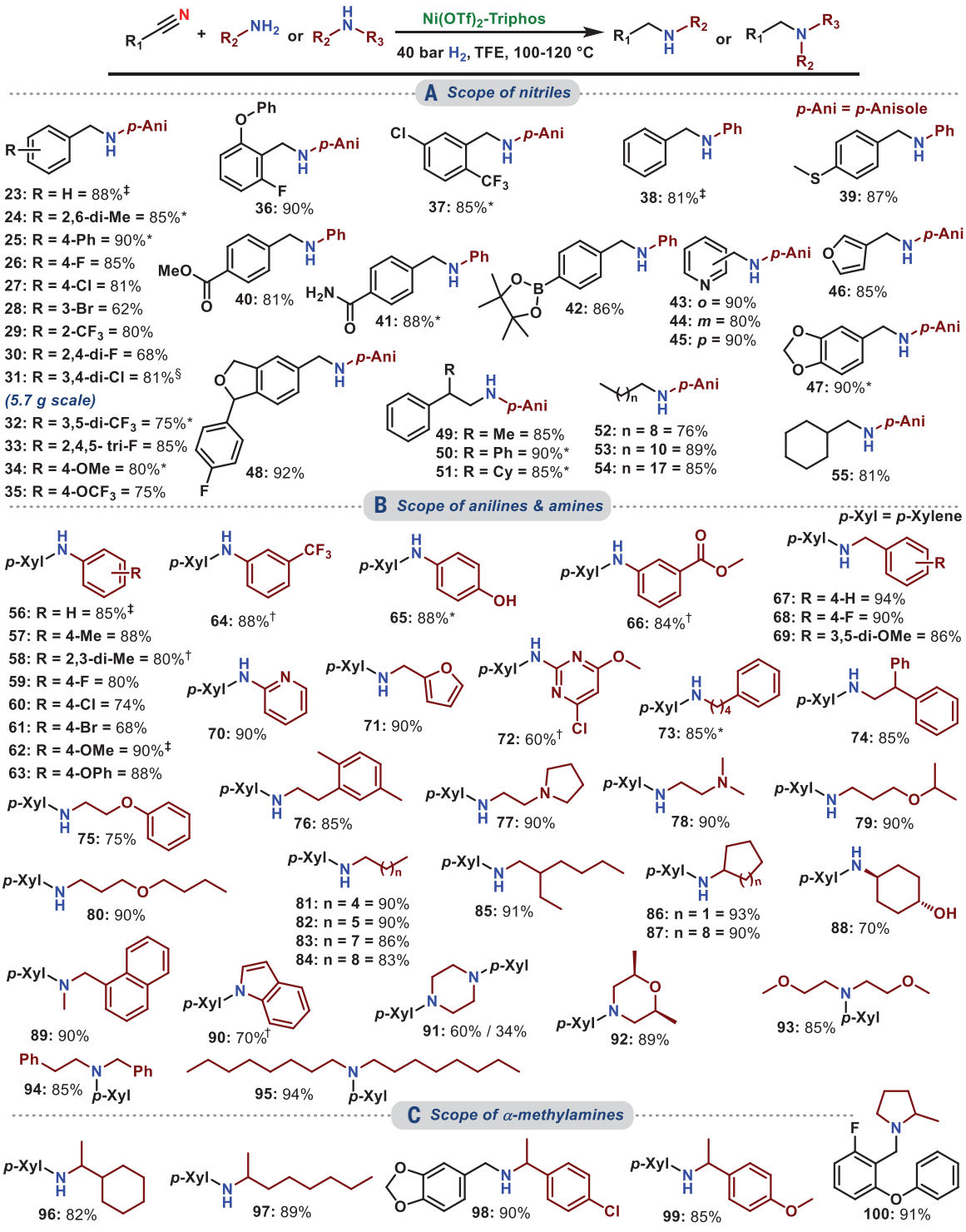


Fig. 3. Synthesis of secondary and tertiary amines by Ni-catalyzed hydrogenative cross coupling of nitriles and amines. Reaction conditions: 0.5 mmol nitrile, 0.75 mmol amine, 4 mol % $Ni(OTf)_2$, 5 mol % linear triphos (L1), 40 bar H_2, 3 mL TFE, 100°C, 16 hours, isolated yields. *, with 5 mol % $Ni(OTf)_2$ and 6 mol % linear triphos (L1) at 120°C; †, with 2 mmol amine; ‡, GC yields with 4-flourotoluene as standard; §, scale up for **31**: 4.3 g 3,4-dichlorobenzonitrile, 4.62 g *p*-anisidine, 446.04 mg (5 mol %) $Ni(OTf)_2$, 801.82 mg (6 mol %) linear triphos, 50 bar H_2, 40 mL TFE, 120°C, 16 hours, isolated yield.

12% of the tertiary amine **22** was observed. However, conducting the reaction in the presence of ammonia gas at 120°C, we obtained 87% of the desired 4-methylbenzylamine **18**. To understand this behavior and to know more about the role of ammonia, we conducted the hydrogenation of *p*-toluonitrile **15** in the presence of ^{15}N-labeled ammonium acetate. Notably, in this case LCMS and NMR analysis showed the predominant formation of the ^{15}N-labeled 4-methylbenzylamine (86% yield; 89 and 11% $^{15}N/^{14}N$-selectivity) (figs. S11 to S13). This result clearly shows that this transformation does not proceed as simple hydrogenation but alternatively as a reductive coupling

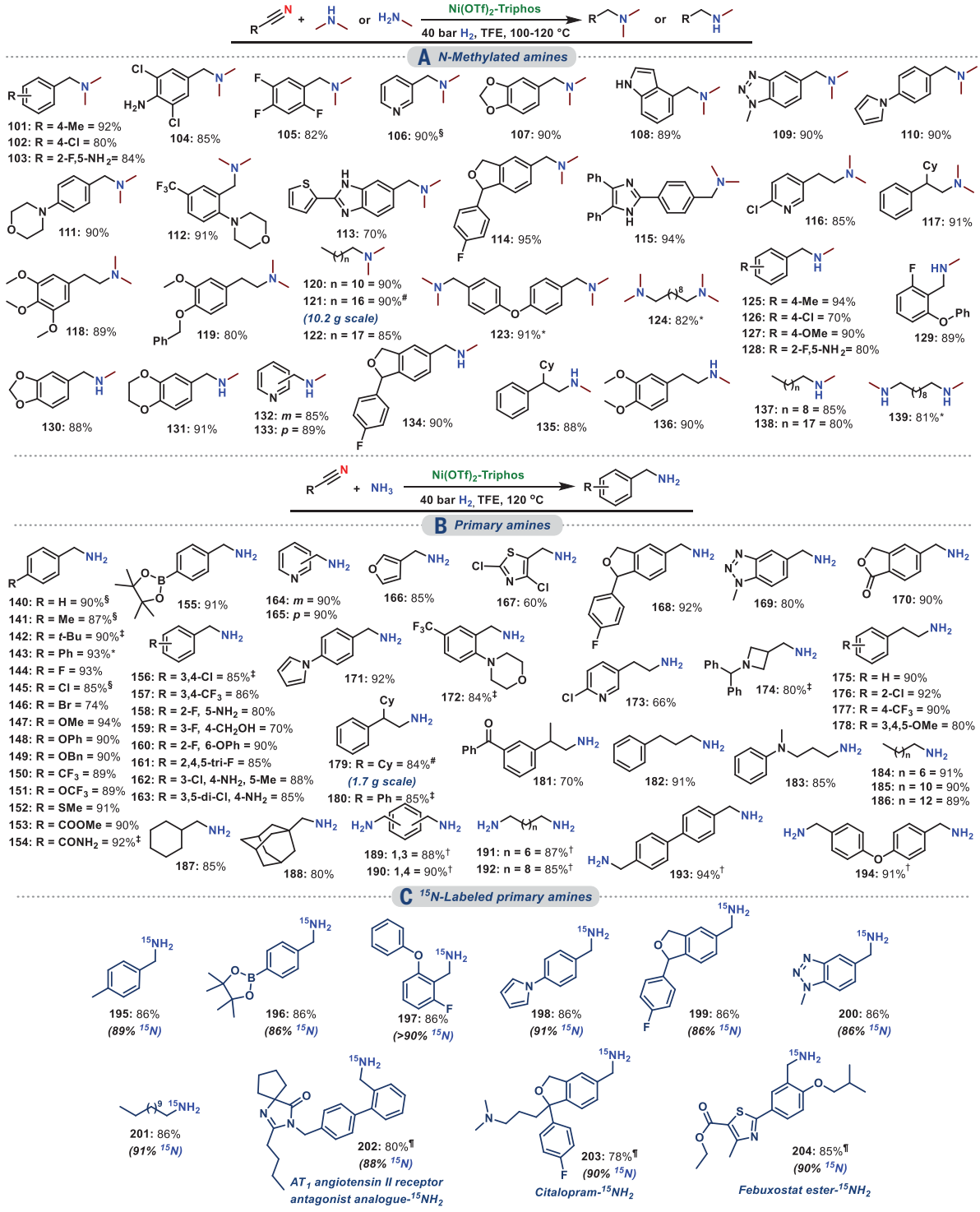


Fig. 4. Ni-catalyzed hydrogenative cross coupling of nitriles with *N*-methylamines and ammonia. Reaction conditions: (**A**) (*N*-methylated amines): 0.5 mmol nitrile, 150 μL *N*,*N*-dimethylamine or *N*-methylamine (12 M in ethanol), 5 mol % $Ni(OTf)_2$, 6 mol % linear triphos (**L1**), 40 bar H_2, 2 mL TFE, 120°C, 8 hours, isolated yields. *, with 7 mol % $Ni(OTf)_2$ and 8 mol % linear triphos (**L1**). (**B**) (Primary amines): 0.5 mmol nitrile, 5 to 7 bar NH_3, 5 mol % $Ni(OTf)_2$, 6 mol % linear triphos (**L1**), 40 bar H_2, 2 mL TFE, 120°C, 5 hours. Isolated as free primary amines and converted to hydrochloride salts for measuring NMR. †, with 7 mol % $Ni(OTf)_2$ and 8 mol % linear triphos (**L1**), 10 hours; ‡, for 10 hours; §, GC yields with 4-flourotoluene as standard. (**C**) (^{15}N-labeled primary amines): 0.5 mmol nitrile, 2.5 mmol $^{15}NH_4OAc$, 5 mol % $Ni(OTf)_2$, 6 mol % linear triphos (**L1**), 40 bar H_2, 2 mL TFE, 120°C, 5 hours. Isolated as free primary amines and converted to hydrochloride salts for measuring NMR. ^{15}N incorporation was determined by ^{1}H-NMR. ¶, with 0.25 mmol nitrile and 1.3 mmol $^{15}NH_4OAc$; #, scale up for **121**: 10.01 g heptadecanonitrile, 12 mL dimethylamine (12 M in EtOH), 713.66 mg (5 mol %) $Ni(OTf)_2$, 1.28 g mg (6 mol %) linear triphos, 50 bar H_2, 50 mL TFE, 120°C, 16 hours. For **179**: 1.99 g 2-cyclohexyl-2-phenylacetonitrile, 5 to 7 bar NH_3, 178.42 mg (5 mol %) $Ni(OTf)_2$, 320.73 mg (6 mol %) linear triphos, 40 bar H_2, 30 mL TFE, 120°C, 10 hours. All are isolated yields.

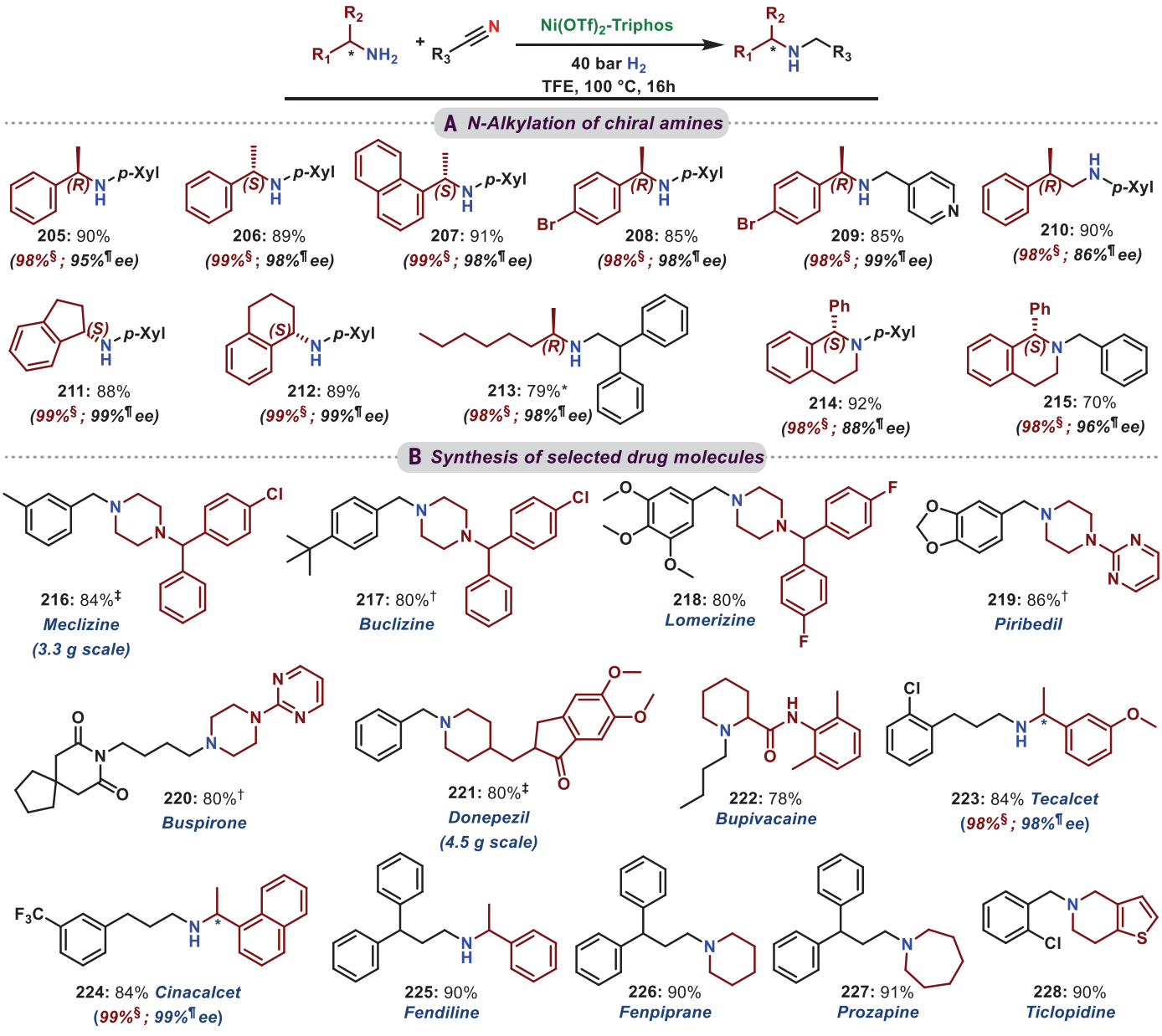


Fig. 5. Applicability of Ni-catalyzed hydrogenative cross-coupling reactions for the *N*-alkylation of chiral amines with nitriles and synthesis of selected drug molecules. Reaction conditions: 0.25 mmol nitrile, 0.37 mmol amine, 4 mol % $Ni(OTf)_2$, 5 mol % linear triphos (**L1**), 40 bar H_2, 3 mL TFE, 100°C, 16 hours, isolated yields. *, at 120°C; †, with 5 mol % $Ni(OTf)_2$ and 6 mol % linear triphos (**L1**), 120°C; ‡, scale up for **216**: 1.17 g 3-methylbenzonitrile, 4.3 g 1-((4-chlorophenyl)(phenyl)methyl)piperazine, 142.73 mg (4 mol %) $Ni(OTf)_2$, 267.27 mg (5 mol %) linear triphos, 40 bar H_2, 30 mL TFE, 100°C, 10 hours, isolated yield. For **221**: 1.55 g benzonitrile, 7.33 g 5,6-dimethoxy-2-(piperidin-4-ylmethyl)-2,3-dihydro-1H-inden-1-one, 214.10 mg (4 mol %) $Ni(OTf)_2$, 401 mg (5 mol %) linear triphos, 40 bar H_2, 50 mL TFE, 100°C, 16 hours, isolated yield. §, ee of starting material (chiral primary amine) obtained commercially; ¶, ee of product.

process (fig. S6). Indeed, increasing the amount of ^{15}N-labeled ammonium acetate in the hydrogenation of **15** correspondingly increased incorporation of ^{15}N in the product: We obtained 89% labeled product by using 5 equiv. of $^{15}NH_4OAc$ (figs. S11 and S12). This result implicates an equilibrium between the ^{15}N-labeled and nonlabeled imine through the corresponding aminal. The selective formation of primary amine in the presence of ammonia requires slightly higher temperatures compared with the cross coupling of the nitrile with aniline. Notably, in the presence of ammonia instead of 4-methoxyaniline the overall reaction is faster (fig. S5) but less selective, and considerable amounts of self-coupling product **21** are obtained at 100°C. However, increasing the temperature to 120°C selectively produced the corresponding primary amine **18**.

Similar to the reactions with primary and secondary amines the Ni-triphos catalyst is both active and selective for the hydrogenative coupling of various nitriles with ammonia to produce corresponding primary amines (Fig. 4B; products **140** to **194**). For example, mono-, di-, and tri-halogen–substituted benzylic amines were efficiently prepared in up to 93% yield (Fig. 4B; products **144** to **146**, **156** to **163**). Again, functional groups such as thioether, ester, amide, and hydroxyl are compatible. Furthermore, different heterocyclic nitriles based on pyridine, benzothiazole, thiazole, benzotriazole, furan, isobenzofuran-1(3H)-one,

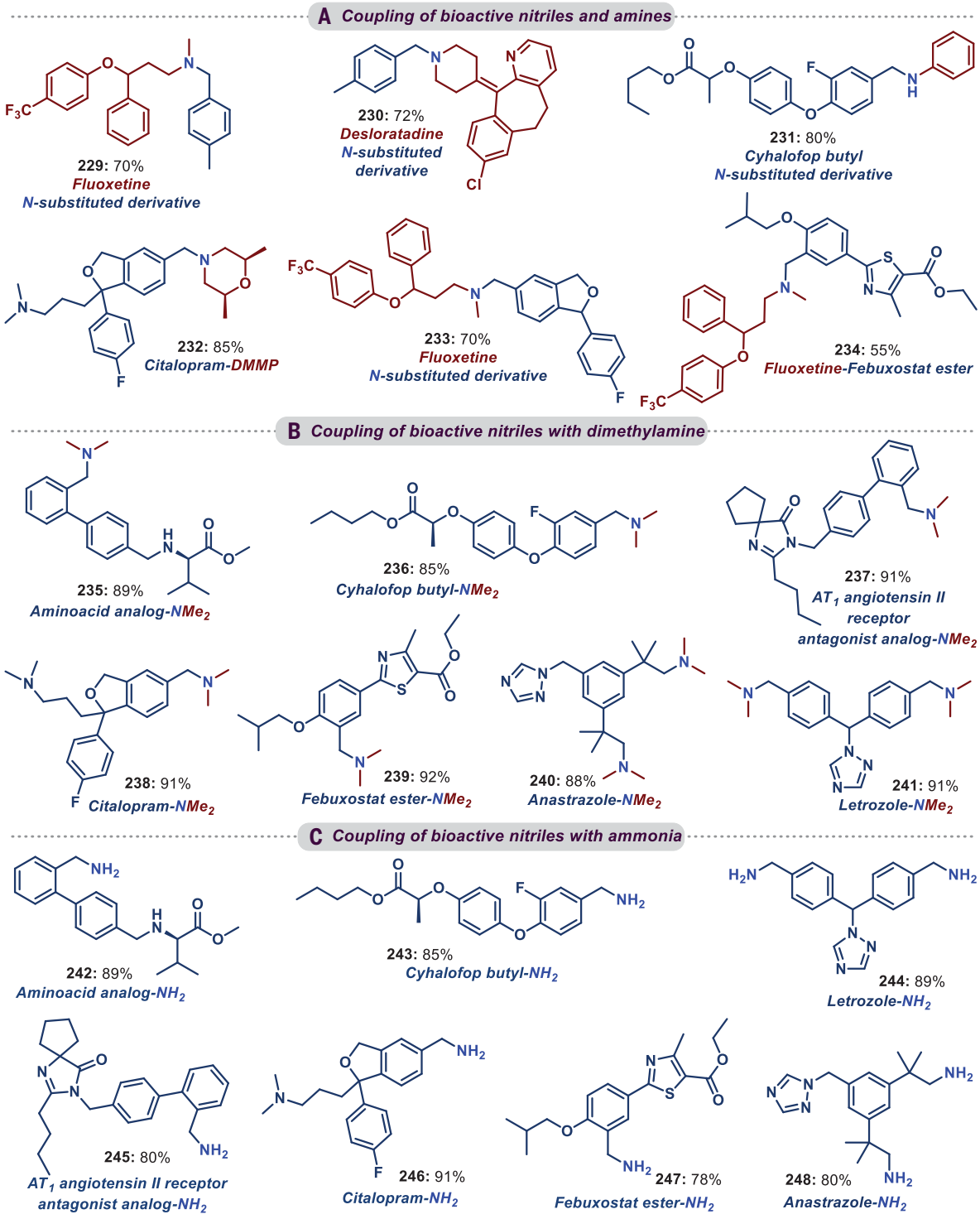


Fig. 6. Late-stage functionalization of biologically active nitriles and amines. Reaction conditions: (**A**) 0.25 mmol nitrile, 0.37 mmol amine, 5 mol % $Ni(OTf)_2$, 6 mol % linear triphos (**L1**), 40 bar H_2, 3 mL TFE, 100°C, 16 hours. (**B**) 0.25 mmol nitrile, 75 μL *N*,*N*-dimethylamine (12 M in ethanol), 5 mol % $Ni(OTf)_2$, 6 mol % linear triphos (**L1**), 40 bar H_2, 2 mL TFE, 120°C, 10 hours. (**C**) 0.25 mmol nitrile, 5-7 bar NH_3, 5 mol % $Ni(OTf)_2$, 6 mol % linear triphos (**L1**), 40 bar H_2, 2 mL TFE, 120°C, 10 hours. All yields are isolated.

1,3-dihydroisobenzofuran, 1-phenyl-1H-pyrrole, 4-phenyl-morpholine, and 2-(thiophen-2-yl)-1H-benzoimidazole smoothly underwent reductive coupling and provided the corresponding heterocyclic primary amines (Fig. 4B; products **164** to **174**). Finally, different phenyl-alkyl, and (cyclo)aliphatic primary amines were successfully prepared with this catalyst system (Fig. 4B; products **175** to **183**). The selective conversion of fatty nitriles to fatty amines represents an important industrial process (*44*, *45*). Indeed, >800,000 tons of fatty amines are produced per annum for the manufacture of fabric softeners, flotation agents, emulsifiers, corrosion inhibitors, and lubricating additives

(*44*). Applying this Ni-triphos catalyst system, different fatty nitriles were coupled with ammonia to the corresponding fatty amines in 88 to 90% yield (Fig. 4B; products **184** to **186** and **191** to **192**).

Likewise, a variety of ^{15}N-labeled primary amines can be prepared from the corresponding nitriles and $^{15}NH_4OAc$. Such transformations are useful for studies probing the metabolism of bioactive compounds and seeking to identify specific metabolites (*46*, *47*). To showcase the utility under optimized conditions 11 ^{15}N-labled primary amines were prepared, including three bioactive molecules. In all these cases up to 91% of ^{15}N incorporation was observed (Fig. 4C; products **195 to 204**).

N-alkylation of chiral amines with nitriles

To highlight the applicability of these catalysts to medicinal and complex organic chemistry, further functionalization of chiral molecules was studied. As shown in Fig. 5A, *N*-alkylation of various chiral primary and secondary amines with different nitriles proceeded smoothly in the presence of molecular hydrogen to achieve *N*-alkylated chiral secondary and tertiary amines with retention of stereoconfiguration. For example, (*R*)-1-phenylethan-1-amine and (*S*)-1-phenylethan-1-amine as well as (*S*)-1-(naphthalen-1-yl)ethan-1-amine reacted selectively with 4-methylbenzonitrile to form the corresponding chiral benzyl amines in up to 90% yields and 98% enantiomeric excess (ee) (Fig. 5A; products **205** to **207**). Similarly, bromo-substituted chiral 1-phenylethan-1-amines gave the desired *N*-alkylated amines with retention of configuration (Fig. 5A; products **208** to **209**). In addition to primary amines, secondary amines led to reductive coupling with nitriles and provided tertiary chiral amines in up to 85% yields and 99% ee (Fig. 5A; products **214** to **215**). Notably, for products **210** and **214**, 12 to 14% racemization is observed which can be explained by the comparably high acidity of the α-hydrogen atom in the case of the starting 2-phenylpropionitril. In the case of product **214** we assume partial racemization by a dehydrogenation or rehydrogenation mechanism due to the stability of the intermediate imine.

Synthesis of drugs and late-stage functionalization of bioactive compounds

Next, the application of this methodology has been showcased for the preparation of 13 existing drug molecules (Fig. 5B; products **216** to **228**). Without further optimizations the corresponding nitriles and secondary or primary amines reacted selectively in the presence of the Ni-triphos catalyst system and molecular hydrogen to give the desired molecules in up to 91% yield. For example, cinacalcet and tecalcet—used to treat secondary hyperparathyroidism, parathyroid carcinoma, and primary hyperparathyroidism—were prepared from chiral primary amines in 84% yield (98 to 99% ee) (Fig. 5B; products **223** and **224**).

Late-stage functionalization represents a useful strategy to introduce specific chemical groups and functionalities in the final steps of the synthesis (*48*, *49*). This process accelerates the preparation of diverse intermediates and chemical libraries for drug discovery (*48*, *49*). In this respect, nitrile-containing pharmaceuticals, agrochemicals, and biomolecules (*26*) can be selectively functionalized to even more valuable amines. Indeed, using our Ni-based protocol the cross coupling of cyhalofop-butyl, citalopram, and amino acid analogs, as well as AT1 angiotensin II receptor antagonist analog, febuxostat ester, letrozole, and anastrazole with amines and ammonia (Fig. 6; products **229** to **248**), proceeded smoothly. In all these cases the -CN groups reacted with amines or ammonia in a highly selective manner without affecting the rest of the core structure. Similarly, derivatives of bioactive amines (e.g., fluoxetine and desloratadine) can be easily alkylated with various nitriles. Finally, to demonstrate the applicability of this homogeneous catalytic protocol the preparation of selected products was scaled up to the 10-gram level without problems (products **31, 121, 179, 216,** and **221**).

In conclusion, a homogenous Ni-triphos catalyst system allows for the hydrogenative cross coupling of various nitriles and amines or ammonia in the presence of molecular hydrogen to produce functionalized and structurally diverse aromatic, heterocyclic, and aliphatic primary, secondary, and tertiary amines including *N*-methylated products. Following this synthetic protocol, nitriles constitute attractive starting materials for the preparation of a broad spectrum of valuable amines including pharmaceutical drugs. In addition, efficient late-stage functionalization of life science molecules is possible. Current limitations of this methodology have been identified and include less nucleophilic amine substrates as well as functional groups such as aldehyde, keto, alkyne, and nitro.

REFERENCES AND NOTES

1. S. A. Lawrence, *Amines: Synthesis, Properties and Applications* (Cambridge Univ. Press, 2004).
2. A. Ricci, *Amino Group Chemistry: From Synthesis to the Life Sciences* (Wiley, 2008).
3. M. H. Qureshi, The Njardarson Group, The University of Arizona, Top 200 Pharmaceuticals by Retail Sales in, 2020 (2020); https://njardarson.lab.arizona.edu/sites/njardarson.lab.arizona.edu/files/Top%20200%20 Pharmaceuticals%20By%20Retail%20Sales%202020V3.pdf.
4. Mordor Intelligence, "Amines Market - Growth, Trends, Covid-19 Impact, and Forecasts (2022-2027)" (2022); https://www.mordorintelligence.com/industry-reports/amines-market.
5. Grand View Research, "Amines Market Size, Share & Trends Analysis Report By Product (Ethanolamines, Fatty Amines, Alkylamines), By Application (Crop Protection, Surfactants, Water Treatment), By Region, And Segment Forecasts, 2018 - 2025" (Report 978-1-68038-291-4, 2022); https://www.grandviewresearch.com/industry-analysis/amines-industry.
6. J. A. Garduño, J. J. García, *ACS Catal.* **10**, 8012–8022 (2020).
7. C. Bornschein *et al*., *Nat. Commun.* **5**, 4111 (2014).
8. K. Murugesan *et al*., *Chem. Sci.* **9**, 8553–8560 (2018).
9. Y. Saito, H. Ishitani, M. Ueno, S. Kobayashi, *ChemistryOpen* **6**, 211–215 (2017).
10. A. Mukherjee, D. Srimani, S. Chakraborty, Y. Ben-David, D. Milstein, *J. Am. Chem. Soc.* **137**, 8888–8891 (2015).
11. D. Formenti, F. Ferretti, F. K. Scharnagl, M. Beller, *Chem. Rev.* **119**, 2611–2680 (2019).
12. R. V. Jagadeesh *et al*., *Science* **342**, 1073–1076 (2013).
13. T. Irrgang, R. Kempe, *Chem. Rev.* **120**, 9583–9674 (2020).
14. O. I. Afanasyev, E. Kuchuk, D. L. Usanov, D. Chusov, *Chem. Rev.* **119**, 11857–11911 (2019).
15. K. Murugesan *et al*., *Chem. Soc. Rev.* **49**, 6273–6328 (2020).
16. T. Yasukawa, R. Masuda, S. Kobayashi, *Nat. Catal.* **2**, 1088–1092 (2019).
17. T. Yan, B. L. Feringa, K. Barta, *Nat. Commun.* **5**, 5602 (2014).
18. T. Irrgang, *Chem. Rev.* **119**, 2524–2549 (2019).
19. P. Ruiz-Castillo, S. L. Buchwald, *Chem. Rev.* **116**, 12564–12649 (2016).
20. J. F. Hartwig, *Acc. Chem. Res.* **41**, 1534–1544 (2008).
21. S. Streiff, F. Jérôme, *Chem. Soc. Rev.* **50**, 1512–1521 (2021).
22. J. Gui *et al*., *Science* **348**, 886–891 (2015).
23. F. F. Fleming, F. F. Fleming, *Nat. Prod. Rep.* **16**, 597–606 (1999).
24. F. F. Fleming, L. Yao, P. C. Ravikumar, L. Funk, B. C. Shook, *J. Med. Chem.* **53**, 7902–7917 (2010).
25. S. Chakraborty, G. Leitus, D. Milstein, *Angew. Chem. Int. Ed.* **56**, 2074–2078 (2017).
26. J. H. Cho, S. Byun, A. Cho, B. M. Kim, *Catal. Sci. Technol.* **10**, 4201–4209 (2020).
27. S. Lu, J. Wang, X. Cao, X. Li, H. Gu, *Chem. Commun.* **50**, 3512–3515 (2014).
28. Z. Shao, S. Fu, M. Wei, S. Zhou, Q. Liu, *Angew. Chem. Int. Ed.* **55**, 14653–14657 (2016).
29. S. K. Sharma *et al*., *Catal. Sci. Technol.* **3**, 85–88 (2013).
30. T. Schönauer, S. L. J. Thomä, L. Kaiser, M. Zobel, R. Kempe, *Chemistry* **27**, 1609–1614 (2021).
31. L. M. Pignolet, *Homogeneous Catalysis with Metal Phosphine Complexes* (Springer, 2013)
32. P. C. J. Kamer, P. W. N. M. van Leeuwen, *Phosphorus(III) Ligands in Homogeneous Catalysis: Design and Synthesis* (Wiley, 2012).
33. Z. Fei, P. J. Dyson, *Coord. Chem. Rev.* **249**, 2056–2074 (2005).
34. T. vom Stein *et al*., *J. Am. Chem. Soc.* **136**, 13217–13225 (2014).
35. Y. Li, C. Topf, X. Cui, K. Junge, M. Beller, *Angew. Chem. Int. Ed.* **54**, 5196–5200 (2015).
36. S. Wesselbaum, T. Vom Stein, J. Klankermayer, W. Leitner, *Angew. Chem. Int. Ed.* **51**, 7499–7502 (2012).
37. J. Kothandaraman, A. Goeppert, M. Czaun, G. A. Olah, G. K. S. Prakash, *J. Am. Chem. Soc.* **138**, 778–781 (2016).
38. J. Schneidewind, R. Adam, W. Baumann, R. Jackstell, M. Beller, *Angew. Chem. Int. Ed.* **56**, 1890–1893 (2017).
39. T. J. Korstanje, J. I. van der Vlugt, C. J. Elsevier, B. de Bruin, *Science* **350**, 298–302 (2015).
40. X. Cui, Y. Li, C. Topf, K. Junge, M. Beller, *Angew. Chem. Int. Ed.* **54**, 10596–10599 (2015).
41. A. A. Núñez Magro, G. R. Eastham, D. J. Cole-Hamilton, *Chem. Commun.* 3154–3156 (2007).
42. J. R. Cabrero-Antonino, R. Adam, V. Papa, M. Beller, *Nat. Commun.* **11**, 3893 (2020).
43. S. S. Mansouri, M. I. Ismail, K. Almoor, "A systematic approach for conceptual and sustainable process design: Production of methylamines from methanol and ammonia" in *Proceedings of the 2012 AIChE Annual Meeting* (American Institute of Chemical Engineers, 2012).
44. A. Hinzmann, H. Gröger, *Eur. J. Lipid Sci. Technol.* **122**, 1900163 (2020).
45. Grand View Research, "Fatty Amines Market Size, Share & Trends Analysis Report By Product (Primary, Secondary, Tertiary), By Application (Agrochemicals, Oilfield Chemicals, Personal Care), And Segment Forecasts, 2019 – 2025" (Report GVR-1-68038-939-5, 2022); https://www.grandviewresearch.com/industry-analysis/fatty-amines-market.
46. R. Schoenheimer, D. Rittenberg, *Science* **87**, 221–226 (1938).
47. J. Buček *et al*., *R. Soc. Open Sci.* **5**, 181322 (2018).
48. T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal, S. W. Krska, *Chem. Soc. Rev.* **45**, 546–576 (2016).
49. M. Moir, J. J. Danon, T. A. Reekie, M. Kassiou, *Expert Opin. Drug Discov.* **14**, 1137–1149 (2019).

ACKNOWLEDGMENTS

We thank the analytical team of the Leibniz-Institut für Katalyse e.V. for their excellent service and A. Spannenberg for x-ray analysis. **Funding:** We gratefully acknowledge the European Research Council (EU project 670986-NoNaCat), German Research Foundation (DFG project 670389-NFDI 2/1) and the State of Mecklenburg-Vorpommern for financial and general support. **Author contributions:** R.V.J. and M.B. supervised the project. R.V.J., V.G.C., and M.B. planned and developed the project. V.G.C., R.V.J., and M.B. designed the experiments. V.G.C. performed catalytic experiments, W.B. and V.G.C. performed ^{15}N NMR analysis, R.V.J., V.G.C., and M.B. wrote the paper. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** X-ray crystal data can be obtained from the joint Cambridge Crystallographic Data Centre (CCDC: 2163405) and Fachinformationszentrum Karlsruhe Access Structures service www.ccdc.cam.ac.uk/structures. All other data remaining are presented in supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn7565
Materials and Methods
Figs. S1 to S13
Tables S1 to S4
References (*50*–*52*)
Crystallography Reports
NMR Data and Spectra

Submitted 17 December 2021; accepted 27 April 2022
10.1126/science.abn7565

NEURODEVELOPMENT

Fate mapping of neural stem cell niches reveals distinct origins of human cortical astrocytes

Denise E. Allen[1,2,3,4], Kevin C. Donohue[2,5,6,7,8], Cathryn R. Cadwell[9], David Shin[1,2,3,4], Matthew G. Keefe[1,2,3,4], Vikaas S. Sohal[2,7,8], Tomasz J. Nowakowski[1,2,3,4,7]*

Progenitors of the developing human neocortex reside in the ventricular and outer subventricular zones (VZ and OSVZ, respectively). However, whether cells derived from these niches have similar developmental fates is unknown. By performing fate mapping in primary human tissue, we demonstrate that astrocytes derived from these niches populate anatomically distinct layers. Cortical plate astrocytes emerge from VZ progenitors and proliferate locally, while putative white matter astrocytes are morphologically heterogeneous and emerge from both VZ and OSVZ progenitors. Furthermore, via single-cell sequencing of morphologically defined astrocyte subtypes using Patch-seq, we identify molecular distinctions between VZ-derived cortical plate astrocytes and OSVZ-derived white matter astrocytes that persist into adulthood. Together, our study highlights a complex role for cell lineage in the diversification of human neocortical astrocytes.

Radial glia serve as the neural stem cells of the developing neocortex (*1*, *2*). During mid-gestation in humans, radial glia can be further classified as either CRYAB$^+$ truncated radial glia (tRG) that lose their pial contact and remain in the ventricular zone (VZ) (*3*), or HOPX$^+$ outer radial glia (oRG) that lose their ventricular contact and reside in the outer subventricular zone (OSVZ) alongside neuronal intermediate progenitor cells (IPCs) (*4*–*10*). These VZ and OSVZ stem cell niches can also be further distinguished on the basis of their extracellular matrix components (*7*, *11*) and signaling pathway activation (*11*). Prior studies have shown that progenitor cells residing in the OSVZ contribute to cortical neurogenesis (*9*, *11*, *12*), astrogliogenesis (*13*), and oligodendrocytes (*14*), but the cells generated by human VZ progenitors are less well characterized (*3*). tRG cells emerge at the onset of astrogliogenesis (*3*), which suggests that they could also contribute to the cortical astrocyte pool. To determine whether both VZ and OSVZ progenitors contribute cortical astrocytes, we performed fate mapping of the human VZ and OSVZ. Our study revealed that VZ and OSVZ progenitors give rise to spatially, morphologically, and molecularly distinct subpopulations of astrocytes, highlighting the complexity of developmental lineage relationships in the developing human brain.

[1]Department of Anatomy, University of California, San Francisco, CA, USA. [2]Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, CA, USA. [3]Department of Neurological Surgery, University of California, San Francisco, CA, USA. [4]Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research, University of California, San Francisco, CA, USA. [5]School of Medicine, University of California, San Francisco, CA, USA. [6]Center for Integrative Neuroscience, University of California, San Francisco, CA, USA. [7]Weill Institute for Neurosciences, University of California, San Francisco, CA, USA. [8]Kavli Institute for Fundamental Neuroscience, University of California, San Francisco, CA, USA. [9]Department of Pathology, University of California, San Francisco, CA, USA.
*Corresponding author. Email: tomasz.nowakowski@ucsf.edu

Cellular output of mid-gestation germinal zones

To fate-map the human VZ and OSVZ niches, we prepared organotypic slice cultures of primary human neocortex from gestational weeks (GW) 18 to 23 to capture the time point shortly after tRG emergence (*3*) and the onset of astrogliogenesis (*13*). We labeled these niches in paired slices from the same individual using local delivery of viral vectors expressing green fluorescent protein (GFP) under the CMV/chicken β-actin (CAG) promoter (Fig. 1A). Two days after labeling—the earliest time point at which the GFP can be detected—GFP$^+$ cells labeled by VZ infections were found in the SOX2^{dense}/CRYAB$^+$ VZ and adjacent inner subventricular zone, whereas cells labeled by OSVZ infections were located in the SOX2^{dense}/CRYAB$^-$ OSVZ with a few cells in the adjacent SOX2^{sparse}/CRYAB$^-$ subplate (SP) (Fig. 1B and fig. S1A). To determine the identity of these initially labeled cells, we stained for several markers of progenitor cell subtypes (Fig. 1, C and F, and fig. S1, B and C). Across slices from three individuals, 76 ± 10% (SD, here and below) of the cells labeled by VZ infection were SOX2$^+$ progenitors, and 10 ± 2% were CRYAB$^+$ tRG (Fig. 1E). After OSVZ labeling, we found that an average of 55 ± 8% of OSVZ-infected GFP$^+$ cells were dividing or newly born cells that incorporated 5-bromo-2′-deoxyuridine (BrdU), 10 ± 4% were EOMES$^+$ neuronal progenitor cells, and 5 ± 2% were HOPX$^+$ oRGs (Fig. 1F). Together, these data show that our local viral labeling is spatially specific and labels progenitor cells.

Next, we compared the distributions of VZ- and OSVZ-derived cells across the VZ, SVZ, SP, and cortical plate (CP) (fig. S2A) after 8 to 12 days in culture (Fig. 1G and fig. S2C). Consistent with prior studies (*9*, *12*, *14*, *15*), both the VZ and OSVZ gave rise to cells with neuronal and glial morphologies (Fig. 1, H to K). However, across four to five individuals, OSVZ-derived glia largely remained within the SVZ, whereas VZ-derived glia were found throughout the cortical wall (Fig. 1L). To confirm that this phenotype was not induced by the serum in our culture media, we repeated this experiment in serum-free conditions and found a similar distribution (fig. S3). To confirm that this observation was not the result of nonspecific labeling during the initial VZ infection, we labeled the VZ using an alternative method of microdissecting out the VZ, separately bathing it in virus, and then co-culturing it with the remaining tissue slice (fig. S4A). After 13 days in culture, the transplanted VZ gave rise to GFP$^+$ cells with both neuronal and glial morphologies (fig. S4, B to I), and the VZ-derived glial cell population was evenly distributed across all laminae (fig. S4J). Together, our data suggest that the VZ and OSVZ give rise to distinct subpopulations of glial cells with distinct localizations within the developing cortex (Fig. 1N).

Differential distribution of astrocytes and oligodendrocytes

Next, we used coimmunostaining to determine whether this differential glial distribution could be attributed to molecularly defined astrocytes (SOX9$^+$/OLIG2$^-$), oligodendrocyte progenitor cells (OPCs, SOX9$^+$/OLIG2$^+$), and/or oligodendrocytes (SOX9$^-$/OLIG2$^+$) (Fig. 2A). We found that all three glial cell types derived from the same niche shared comparable distribution patterns (Fig. 2B), indicating that this differential glial output is a broad feature of mid-gestation gliogenesis. These distribution patterns were also different from those of interneurons (fig. S5C), excitatory neurons (fig. S5C), or all cells (fig. S5E). However, we also found an enrichment of astrocytes in the cortical plate of VZ-labeled slices compared to OSVZ-labeled that did not apply to OPCs or oligodendrocytes (Fig. 2C and fig. S5D). Thus, although the differential glial output of VZ and OSVZ niches extends to all macroglial cell types, astrocytes showed a unique differential enrichment in the CP compared to OPCs and oliodendrocytes (Fig. 2D).

To further confirm that the lack of OSVZ-derived glial migration was not due to a broader disruption of normal radial migration in our slice culture model, we determined the distribution of neuronal subtypes by staining for the glutamatergic neuron markers NEUROD2 and TBR1 and the GABAergic neuron marker DLX2 (fig. S5A). We found that DLX2$^+$ cortical interneurons composed a relatively small proportion of GFP$^+$ cells in both VZ and OSVZ infections (fig. S5B) but that they successfully

migrated radially from their niche of origin (fig. S5C). Culture in the presence of BrdU revealed that BrdU$^+$ GFP$^+$/DLX2$^+$ interneurons derived from both VZ and OSVZ labeling (fig. S5F), consistent with their dual origin in cortical and subcortical germinal zones in humans (*16*, *17*). By contrast, the majority of GFP$^+$ cells derived from VZ or OSVZ labeling were NEUROD2$^+$ glutamatergic neurons (fig. S5B) that migrated radially outward from their niche of origin (fig. S5C) and included both BrdU$^-$ and newborn BrdU$^+$ GFP$^+$/TBR1$^+$ glutamatergic neurons

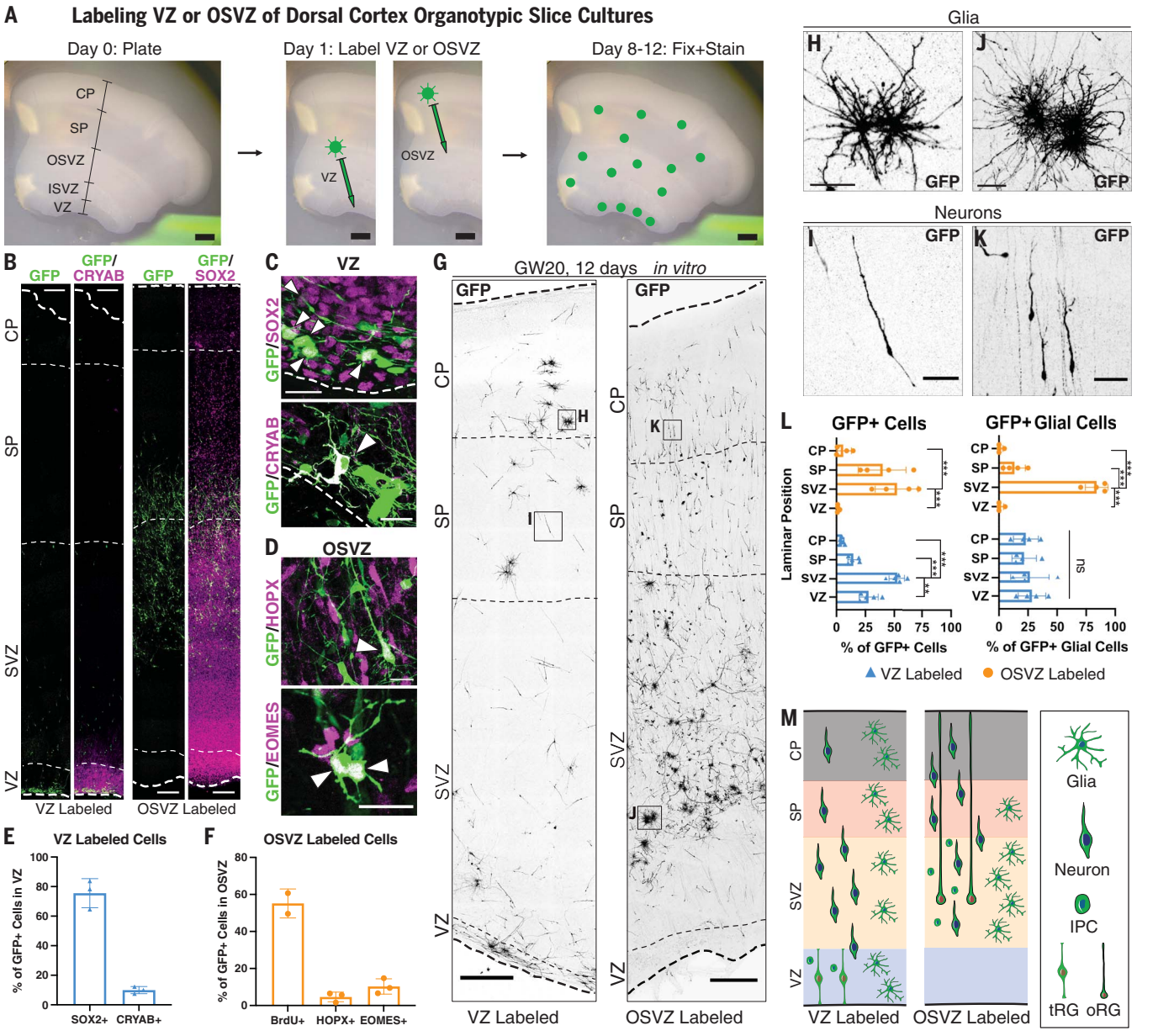


Fig. 1. VZ and OSVZ niches both give rise to neurons but contribute distinct populations of glia. (**A**) Schematic of experimental design. VZ and OSVZ are visualized on a dissection microscope, virus is applied to the VZ or OSVZ using a glass needle, and after 8 to 12 days, labeled GFP$^+$ cells (green dots) have migrated throughout the slice. Scale bar, 1 mm. VZ, ventricular zone; ISVZ, inner subventricular zone; OSVZ, outer subventricular zone; SP, subplate; CP, cortical plate. (**B**) GW19 sample 2 days after VZ or OSVZ labeling coimmunostained for GFP and CRYAB or SOX2. Thick dashed lines indicate ventricular and pial edges; thin dashed lines indicate borders between laminae. Scale bars, 250 μm. (**C**) GFP$^+$/SOX2$^+$ or GFP$^+$/CRYAB$^+$ cells in the VZ 2 days after VZ labeling. White arrowheads indicate double-positive cells. Scale bar, 25 μm. (**D**) GFP$^+$/HOPX$^+$ or GFP$^+$/EOMES$^+$ cells in the OSVZ 2 days after OSVZ labeling. White arrowheads indicate double-positive cells. Scale bar, 25 μm. (**E**) Quantification (mean ± SD) of the percentage of GFP$^+$ cells in the VZ that are SOX2$^+$ or CRYAB$^+$ 2 days after VZ labeling. N = 3 independent samples. (**F**) Quantification (mean ± SD) of the percentage of GFP$^+$ cells in the OSVZ that are BrdU$^+$, HOPX$^+$, or EOMES$^+$ 2 days after OSVZ labeling. N = 2 or 3 independent samples. (**G**) GFP immunostaining of GW20 sample 12 days after VZ or OSVZ labeling. Dashed lines as in (B). Scale bar, 500 μm. (**H** to **K**) Insets from (G) showing cells with glial morphology [(H) and (J)] and neuronal morphology [(I) and (K)]. Scale bars, 50 μm. (**L**) Quantification (mean ± SD) of the percentage of total VZ- or OSVZ-derived GFP$^+$ cells or GFP$^+$ cells with glial morphology located in each lamina of the developing cortex at the end of culture (n = 4 or 5 independent samples). Laminar borders were determined using DAPI density (fig. S2A). *q < 0.033, **q < 0.002, ***q < 0.001 [one-way analysis of variance (ANOVA) with Tukey's multiple-comparisons correction]; ns, not significant. (**M**) Schematic summarizing distribution of VZ- or OSVZ-derived cells.

(fig. S5G). Taken together, these findings demonstrate that extensive neuronal migration in our slice cultures are consistent with the expected distributions, indicating that the slice culture model can support normal patterns of cellular migration.

Morphological analysis of glial cell types

Radial glia and astrocytes undergo morphological transitions across their maturation trajectory, and morphology is a critical feature used to identify astrocyte subtypes in the absence of robust molecular markers (*18*). To determine whether VZ- and OSVZ-derived glia are morphologically heterogeneous, we analyzed the morphology of 524 randomly selected cells with glial morphology across four replicates of paired VZ- and OSVZ-labeled slices from GW20 to GW23. We classified these cells into "morphotypes" according to three previously described criteria: (i) the length of their primary processes, (ii) the density of their primary processes, and (iii) the types of varicosities along their processes (table S1). This morphometric analysis resulted in the identification of 12 glial morphotypes (Fig. 3A and fig. S6) that are consistent with previous morphology annotations at similar ages (*18*–*20*).

We observed several expected radial glial morphotypes including radial glia with a single apical process (*21*) (Fig. 3A and fig. S6: "radial glia—no radial processes") that were derived from and resided in the VZ of younger samples (fig. S7, A to C), as well as radial glia at various stages of transforming into astrocytes (Fig. 3A and fig. S6: "radial glia—short radial processes" and "radial glia—long radial processes") and oRG-like cells with both apical and basal fibers (Fig. 3A and fig. S6: "radial glia—two long processes") derived from and residing in the SVZ of older samples (fig. S7, A to C) (*12*). We also observed SOX9$^+$/OLIG2$^+$/PDGFRA$^+$ OPCs with few, smooth processes (Fig. 3A and figs. S6 and S8A: "sparse smooth processes") (*22*), and SOX9$^-$/OLIG2$^+$/SOX10$^+$ immature oligodendrocytes with many short, branched processes (Fig. 3A and figs. S6 and S8B: "bushy processes") (*22*). A subset of cells with "bushy processes" were also beginning to express the mature oligodendrocyte marker MBP (fig. S8C). These putative oligodendrocyte lineage cells were derived from both the VZ and OSVZ and were found across the cortical wall of all ages studied (fig. S7, A to C).

Lineage divergence of morphologically distinct astrocytes

The remaining morphotypes that did not possess a primary fiber and did not express OLIG2 represent our putative astrocyte subtypes. These morphotypes included several categories with a sparser arbor of primary processes: those with "short processes," those with "sparse beady processes" defined by long thin processes with small, regularly sized and spaced swellings, and those with "sparse processes with irregular varicosities" defined by larger and more irregularly sized and spaced varicosities (Fig. 3A and fig. S6). These morphotypes were derived from both VZ and OSVZ labeling and were found in multiple layers of the cortical wall (fig. S7, B and C).

Cells with a dense arbor of primary processes fell into three subcategories: "dense bulbous processes" with large bulbous varicosities along their processses, "dense smooth processes" with few varicosities, and "dense beady processes" with regularly spaced beady varicosities (Fig. 3A and fig. S6). The "dense bulbous processes" and "dense smooth processes" morphotypes were unique in that they demonstrated a bias toward a single niche of origin and laminar localization. Across four individuals, 82 ± 13% of astrocytes with "dense smooth processes" were derived from OSVZ labeling (Fig. 3B) and 78 ± 10% remained in the SVZ (Fig. 3C) where they represented the dominant astrocyte morphotype (Fig. 3D and fig. S7D). Gamma retrovirus labeling (fig. S10) and BrdU incorporation experiments (Fig. 3E) demonstrated that the majority (70 ± 13%) of these cells were born during the culture period (fig. S7E), indicating that their lack

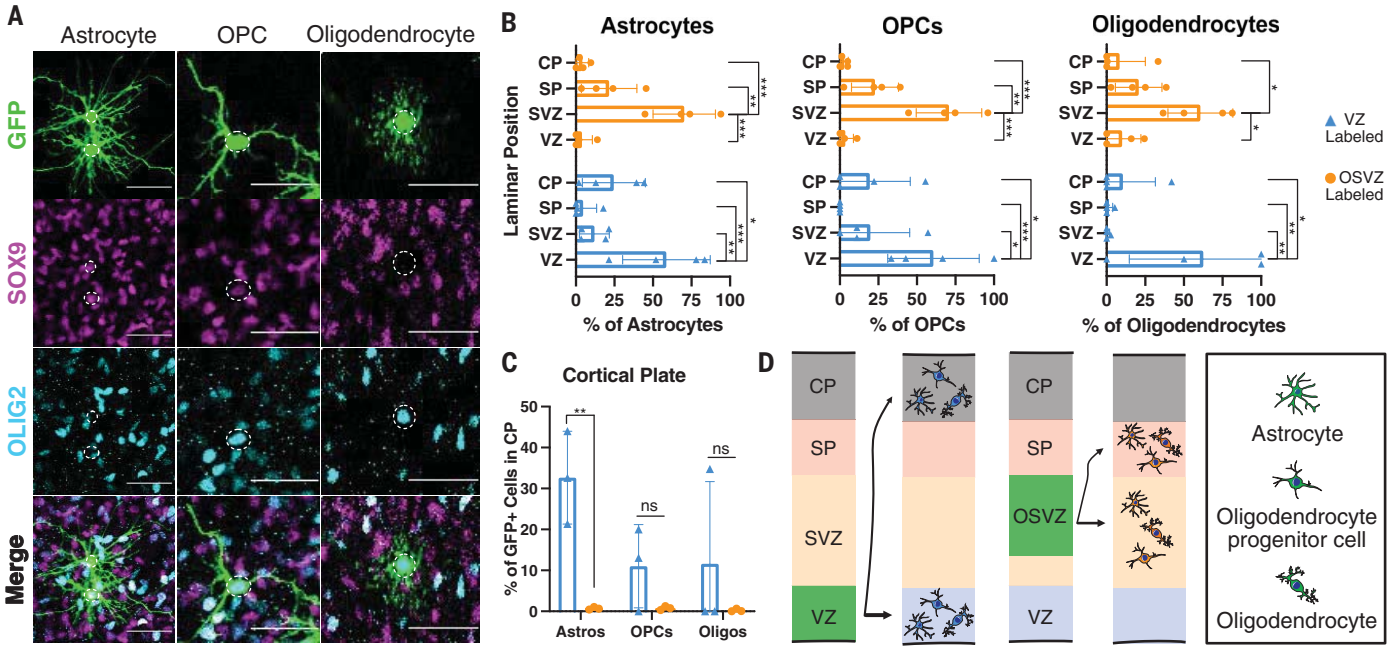


Fig. 2. VZ and OSVZ give rise to spatially distinct populations of astrocytes, OPCs, and oligodendrocytes. (**A**) Coimmunostaining showing representative GFP$^+$/SOX9$^+$/OLIG2$^-$ astrocytes, GFP$^+$/SOX9$^+$/OLIG2$^+$ OPC, and GFP$^+$/SOX9$^-$/OLIG2$^+$ oligodendrocyte. All scale bars, 50 μm. (**B**) Quantification (mean ± SD) of the percentage of VZ- or OSVZ-derived astrocytes, OPCs, or oligodendrocytes that were located in the VZ, SVZ, subplate, or cortical plate at the end of culture. N = 4 independent samples for all. (**C**) Quantification (mean ± SD) of the percentage of total GFP$^+$ cells in the cortical plate of VZ- or OSVZ-labeled slices that were astrocytes (Astros), OPCs, or oligodendrocytes (Oligos). N = 3 independent samples. *q < 0.033, **q < 0.002, ***q < 0.001 (two-way ANOVA with Tukey's multiple-comparisons correction). (**D**) Schematic summarizing the distinct distribution of VZ- versus OSVZ-derived astrocytes, OPCs, and oligodendrocytes. Green color indicates the niche of origin. Arrow line weight corresponds to relative abundance in the target layer.

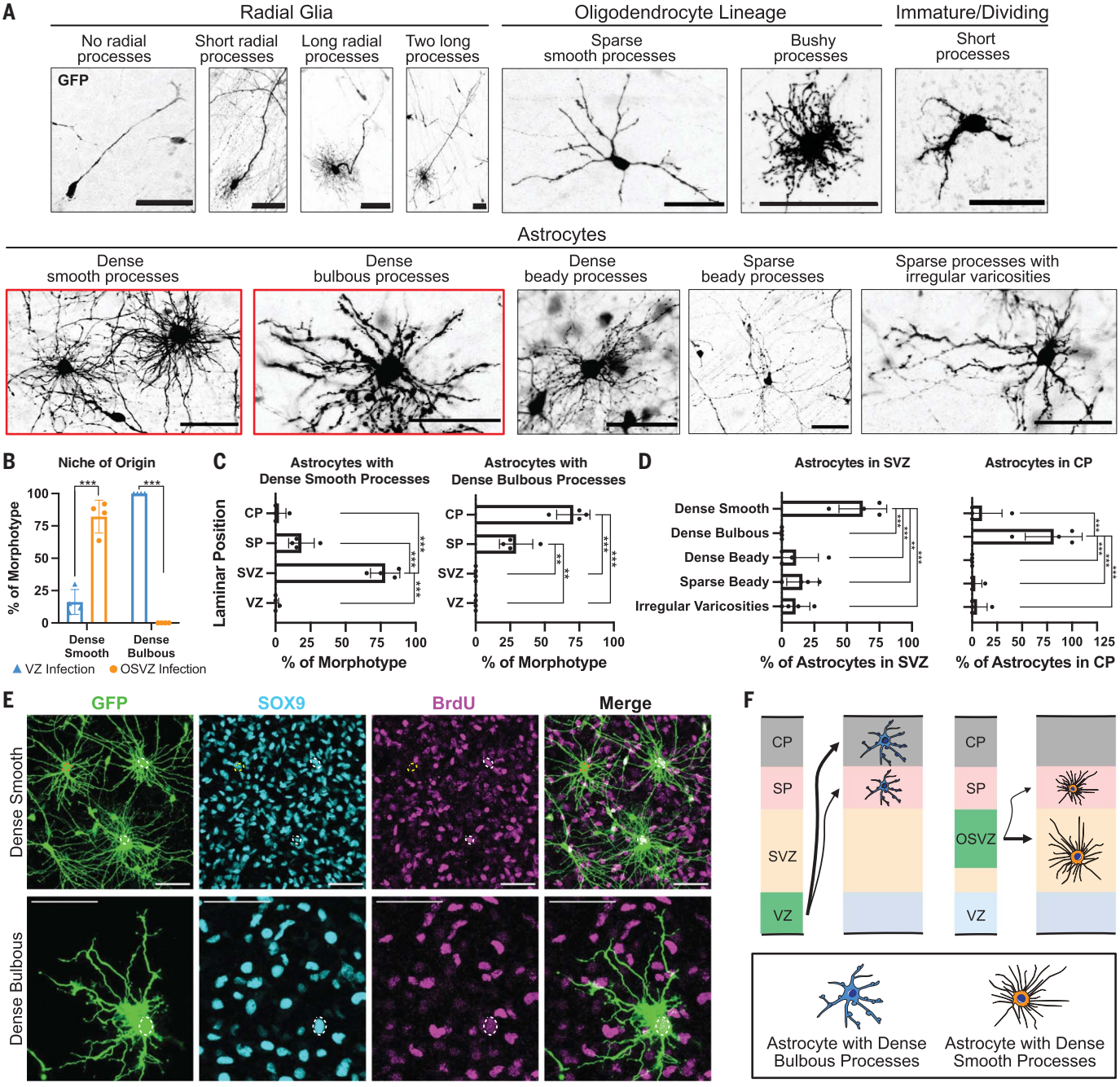


Fig. 3. VZ and OSVZ give rise to morphologically distinct astrocyte subtypes. (**A**) GFP immunostaining showing representative images of the 12 glial morphotypes identified in this study. Scale bars, 50 μm. Red boxes outline morphotypes analyzed later in the figure. (**B**) Quantification of the percentage of astrocytes with dense smooth or dense bulbous processes derived from VZ or OSVZ labeling. (**C**) Quantification of the percentage of astrocytes with dense smooth or dense bulbous processes located in each lamina at the end of culture. (**D**) Quantification of the percentage of GFP^+ astrocytes within the SVZ or CP that adopted each of our five identified astrocyte morphologies. In (B) to (D), $N = 4$ independent pairs of VZ and OSVZ infections; data are means ± SD. $*q < 0.033$, $**q < 0.002$, $***q < 0.001$ (one- or two-way ANOVA with Tukey's multiple-comparisons correction). (**E**) Representative images of $GFP^+/SOX9^+$ "dense smooth" or "dense bulbous" astrocytes that are $BrdU^+$ (white dashed outline) or $BrdU^-$ (yellow dashed outline). Scale bar, 50 μm. (**F**) Schematic summarizing the differential contribution of the VZ and OSVZ to astrocyte morphotypes. Arrow line weight corresponds to relative contribution to the target layer of the cortical wall.

of migration out of the SVZ was not due to labeling of mature cells. By contrast, 100% of "dense bulbous" astrocytes were derived from VZ labeling, and BrdU incorporation confirmed that they were born in this niche (Fig. 3B). The majority of these cells (70 ± 12%) were located in the cortical plate (Fig. 3C) where they also composed the dominant morphotype (Fig. 3D), in addition to 29 ± 12% in the subplate (Fig. 3C). A subset of cells of this morphotype expressed AQP4 consistent with their astrocytic identity, and a subset were positive for Ki-67 (fig. S9), indicating that they could act as a locally dividing progenitor population (*23*). Finally, we confirmed

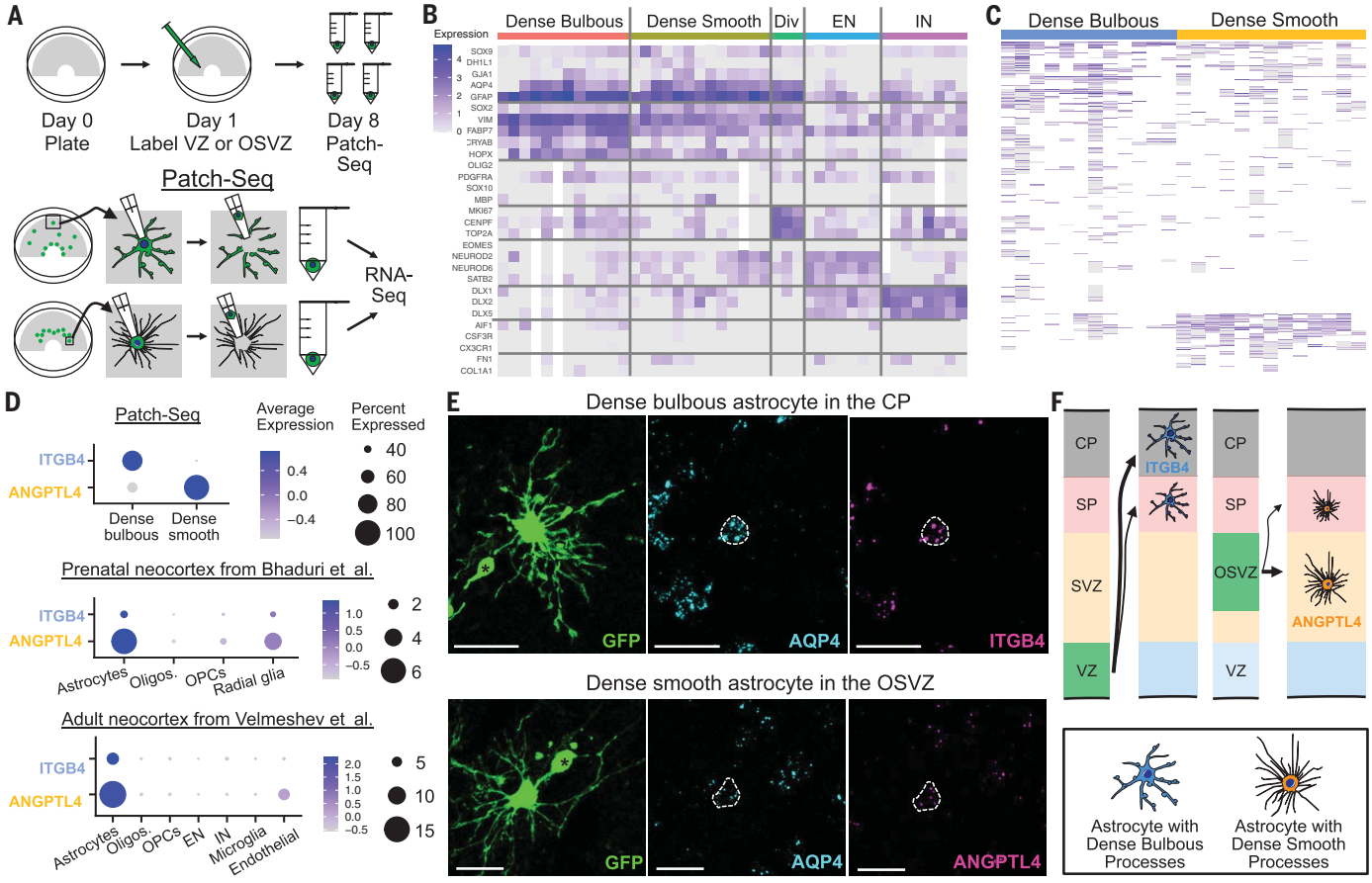


Fig. 4. "Dense bulbous" and "dense smooth" astrocytes have unique molecular profiles. (**A**) Schematic depicting the collection of mRNA from morphologically defined astrocyte morphotypes using Patch-seq. (**B**) Heatmap depicting expression of major cell type marker genes across the 45 cells collected by Patch-seq and used for further analysis. Horizontal gray lines indicate the borders between expression of markers of astrocytes, radial glia, oligodendrocyte lineage cells, dividing cells, excitatory neurons, inhibitory neurons, microglia, and endothelial cells. Div, dividing cells; EN, excitatory neurons; IN, inhibitory neurons. (**C**) Heatmap depicting expression of the 445 genes differentially expressed between "dense bulbous" and "dense smooth" astrocytes (rows), plotted across the 25 astrocytes collected by Patch-seq (columns). (**D**) Dot plots depicting the expression of ITGB4 and ANGPTL4 across astrocyte morphotypes collected by Patch-seq and across cell types in the prenatal or adult cortex. Oligos, oligodendrocytes; OPCs, oligodendrocyte progenitor cells. (**E**) In situ hybridization demonstrating expression of ITGB4 in a VZ-derived dense bulbous astrocyte in the CP and ANGPTL4 expression in an OSVZ-derived dense smooth astrocyte in the OSVZ. (**F**) Schematic summarizing the contribution of VZ and OSVZ to ITGB4$^+$ "dense bulbous" astrocytes and ANGPTL4$^+$ "dense smooth" astrocytes.

that neither of these morphotypes were induced by our culture conditions by replicating our findings in serum-free media (fig. S3). Together, this analysis demonstrates that during mid-gestation the VZ and OSVZ give rise to astrocytes with distinct morphologies and laminar positions within the developing cerebral cortex (Fig. 3F).

Molecular characterization of astrocyte morphotypes

Finally, we sought to determine the molecular distinctions between VZ-derived "dense bulbous" and OSVZ-derived "dense smooth" astrocytes. To make this comparison, we used a modified version of the recently published Patch-seq protocol (*22*) to use a micropipette to collect the mRNA from morphologically defined cells followed by whole-transcriptome single-cell sequencing (Fig. 4A and fig. S11A). Across 10 individuals, we generated Patch-seq data for 28 "dense bulbous" astrocytes, 20 "dense smooth" astrocytes, and 22 neurons to serve as an outgroup. After quality control filtering (see supplementary materials) (fig. S11, B to D), we refined our dataset down to 12 dense bulbous astrocytes, 13 dense smooth astrocytes, three dividing astrocytes, six excitatory neurons, and eight inhibitory neurons of high quality (Fig. 4B). We then performed differential gene expression analysis to identify 445 genes differentially enriched in "dense bulbous" astrocytes or "dense smooth" astrocytes (Fig. 4C and fig. S11E). To identify the genes that most likely mark these cell types across their lifespan, we cross-referenced our differentially expressed gene list with their expression in published single-cell RNA sequencing from either the developing (*24*) or adult (*25*) human neocortex (fig. S11, F and G). This analysis highlighted ITGB4 and ANGPTL4 as candidate markers of dense bulbous and dense smooth astrocytes, respectively, that were also enriched in astrocytes in both prenatal and adult cortex (Fig. 4D). Finally, we confirmed this prediction by performing in situ hybridization on cultured VZ- or OSVZ-labeled slices. This work demonstrated that *ITGB4* mRNA indeed localizes to VZ-derived dense bulbous astrocytes in the cortical plate, and *ANGPTL4* mRNA localizes to OSVZ-derived dense smooth astrocytes in the OSVZ (Fig. 4E and fig. S12, A to C). Together, these experiments demonstrate that the two neural stem cell niches of the developing human cerebral cortex give rise to spatially, morphologically, and molecularly unique astrocyte subtypes (Fig. 4H).

Discussion

Our study illuminates early developmental events that underlie the emergence of human neocortical astrocyte diversity. We find that during mid-gestation, OSVZ progenitor cells give rise to white matter astrocytes, while VZ progenitors give rise to more superficial gray matter astrocytes. This differentiation pattern differs from neurogenesis, where OSVZ progenitors give rise to neurons destined for the upper cortical layers. Because neurogenesis and gliogenesis overlap extensively in humans (*26*), unlike in rodents (*27*), further studies will be needed to determine if individual human progenitors can give rise to both neuronal and glial cells during this time period or follow a classical neurogenic to gliogenic "switch," and whether these patterns of differentiation are maintained throughout development. Our study also implicates truncated radial glia (*3*), which have not been described in mice, in the generation of a distinct subtype of cortical "dense bulbous" astrocytes that have been previously observed in humans (*20*, *28*), but not in rodents. Additional studies are needed to determine the developmental origin of other astrocyte subtypes thought to be enriched in humans, such as varicose projection and interlaminar astrocytes (*29*, *30*).

Astrocytes have been increasingly implicated in neurological disease (*31*, *32*), with distinct astrocyte subtypes being implicated in schizophrenia, bipolar disorder (*33*), and autism spectrum disorders (*25*). Molecular distinctions between white matter and gray matter astrocytes identified in this study may enable insights into their unique developmental features and disease implications. For example, many genes enriched in "dense bulbous" astrocytes have been implicated in glioblastoma cell proliferation and invasion, such as *ITGB4* (*34*), *TMEM158* (*35*), *MGMT* (*36*), and *CELSR1* (*37*), which could explain their remarkable migratory behavior. Detailed mapping of the developmental trajectories of these early astrocyte subtypes through the third trimester and into adulthood will be important to determine whether this astrocyte diversity is encoded at the level of progenitors or dynamically influenced by environmental cues present in distinct brain regions or cellular neighborhoods. In turn, these studies may reveal important subtype-specific vulnerabilities to environmental or genetic insults that underlie this diverse role for astrocytes in neurological disease.

REFERENCES AND NOTES

1. C. R. Cadwell, A. Bhaduri, M. A. Mostajo-Radji, M. G. Keefe, T. J. Nowakowski, *Neuron* **103**, 980–1004 (2019).
2. B. Libé-Philippot, P. Vanderhaeghen, *Annu. Rev. Genet.* **55**, 555–581 (2021).
3. T. J. Nowakowski, A. A. Pollen, C. Sandoval-Espinosa, A. R. Kriegstein, *Neuron* **91**, 1219–1227 (2016).
4. E. Taverna, M. Götz, W. B. Huttner, *Annu. Rev. Cell Dev. Biol.* **30**, 465–502 (2014).
5. N. Kalebic *et al*., *Cell Stem Cell* **24**, 535–550.e9 (2019).
6. N. Kalebic, W. B. Huttner, *Trends Neurosci.* **43**, 843–853 (2020).
7. S. A. Fietz *et al*., *Nat. Neurosci.* **13**, 690–699 (2010).
8. I. H. M. Smart, C. Dehay, P. Giroud, M. Berland, H. Kennedy, *Cereb. Cortex* **12**, 37–53 (2002).
9. D. V. Hansen, J. H. Lui, P. R. L. Parker, A. R. Kriegstein, *Nature* **464**, 554–561 (2010).
10. I. Reillo, C. de Juan Romero, M. Á. García-Cabezas, V. Borrell, *Cereb. Cortex* **21**, 1674–1694 (2011).
11. A. A. Pollen *et al*., *Cell* **163**, 55–67 (2015).
12. M. Betizeau *et al*., *Neuron* **80**, 442–457 (2013).
13. L. C. deAzevedo *et al*., *J. Neurobiol.* **55**, 288–298 (2003).
14. B. G. Rash *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7089–7094 (2019).
15. W. Huang *et al*., *Cell* **182**, 594–608.e11 (2020).
16. O. Marín, J. L. Rubenstein, *Nat. Rev. Neurosci.* **2**, 780–790 (2001).
17. R. N. Delgado *et al*., *Nature* **601**, 397–403 (2022).
18. M. Marín-Padilla, *J. Comp. Neurol.* **357**, 554–572 (1995).
19. C. Falcone *et al*., *Cereb. Cortex* **31**, 379–395 (2021).
20. H. Akiyama, I. Tooyama, T. Kawamata, K. Ikeda, P. L. McGeer, *Brain Res.* **632**, 249–259 (1993).
21. D. E. Schmechel, P. Rakic, *Anat. Embryol.* **156**, 115–152 (1979).
22. D. E. Bergles, W. D. Richardson, *Cold Spring Harb. Perspect. Biol.* **8**, a020453 (2015).
23. W.-P. Ge, A. Miyawaki, F. H. Gage, Y. N. Jan, L. Y. Jan, *Nature* **484**, 376–380 (2012).
24. A. Bhaduri *et al*., *Nature* **598**, 200–204 (2021).
25. D. Velmeshev *et al*., *Science* **364**, 685–689 (2019).
26. S. Malik *et al*., *J. Neurosci.* **33**, 411–423 (2013).
27. S. C. Noctor, V. Martínez-Cerdeño, L. Ivic, A. R. Kriegstein, *Nat. Neurosci.* **7**, 136–144 (2004).
28. V. Matyash, H. Kettenmann, *Brain Res. Rev.* **63**, 2–10 (2010).
29. C. Falcone *et al*., *Glia* **70**, 145–154 (2022).
30. C. Falcone *et al*., *J. Comp. Neurol.* **527**, 1654–1674 (2019).
31. H.-G. Lee, M. A. Wheeler, F. J. Quintana, *Nat. Rev. Drug Discov.* **21**, 339–358 (2022).
32. K. A. Guttenplan *et al*., *Nature* **599**, 102–107 (2021).
33. C. Hercher, V. Chopra, C. L. Beasley, *J. Psychiatry Neurosci.* **39**, 376–385 (2014).
34. Y. Hu *et al*., *Neoplasia* **14**, 219–227 (2012).
35. J. Li *et al*., *Cancer Gene Ther.* 10.1038/s41417-021-00414-5 (2022).
36. M. E. Hegi *et al*., *N. Engl. J. Med.* **352**, 997–1003 (2005).
37. G. Wang *et al*., *Cancer Manag. Res.* **12**, 8857–8865 (2020).

ACKNOWLEDGMENTS

We thank the whole Nowakowski lab for their thoughtful feedback on this project and paper, D. Bogdanoff for assistance in optimizing our Patch-seq protocol, and A. Alvarez-Buylla for helpful feedback on the manuscript. **Funding:** Supported by NIH grants R01NS123263 and U01MH115747, Simons Foundation grant SFARI 491371, an Innovation Award from the Broad Foundation, a New Frontiers Research Award from the Sandler Program for Breakthrough Biomedical Research, and gifts from Schmidt Futures, Shurl and Kay Curci Foundation, Sontag Foundation, and William K. Bowes Jr. Foundation (T.J.N.); NIH grants R25 NS070680 and K08NS126573 (C.R.C.); the UCSF Biomedical Sciences PhD Program and the NSF Graduate Research Fellowships Program (D.E.A. and D.S.); the Ralph H. Kellogg Endowed Chancellor's Fellowship (D.E.A.); and a DoD NDSEG Fellowship (M.G.K.). **Author contributions:** Conceptualization: D.E.A., T.J.N. Methodology: D.E.A., K.C.D., C.R.C., D.S., M.G.K., T.J.N. Investigation: D.E.A., K.C.D., C.R.C., D.S., M.G.K. Visualization: D.E.A., M.G.K., T.J.N. Funding acquisition: D.E.A., K.C.D., D.S., V.S.S., T.J.N. Project administration: D.E.A., T.J.N. Supervision: V.S.S., T.J.N. Writing–original draft: D.E.A., T.J.N. Writing–review and editing: D.E.A., K.C.D., C.R.C., D.S., M.G.K., T.J.N. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Morphology classification data are provided in table S1. Patch-seq counts are provided in table S2 and corresponding metadata are provided in table S3. Code used to generate the data is available at the Github repository: https://zenodo.org/badge/latestdoi/471078301. Materials requests should be addressed to T.J.N. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm5224
Materials and Methods
Figs. S1 to S12
Tables S1 to S3
References (*38*–*42*)
MDAR Reproducibility Checklist

Submitted 23 September 2021; accepted 10 May 2022
Published online 19 May 2022
10.1126/science.abm5224

GEOMORPHOLOGY

Rapid changes to global river suspended sediment flux by humans

Evan N. Dethier[1,2]*, Carl E. Renshaw[1], Francis J. Magilligan[3]

Rivers support indispensable ecological functions and human health and infrastructure. Yet limited river sampling hinders our understanding of consequential changes to river systems. Satellite-based estimates of suspended sediment concentration and flux for 414 major rivers reveal widespread global change that is directly attributable to human activity in the past half-century. Sediment trapping by dams in the global hydrologic north has contributed to global sediment flux declines to 49% of pre-dam conditions. Recently, intensive land-use change in the global hydrologic south has increased erosion, with river suspended sediment concentration on average 41 ± 7% greater than in the 1980s. This north-south divergence has rapidly reconfigured global patterns in sediment flux to the oceans, with the dominant sources of suspended sediment shifting from Asia to South America.

River basins globally are experiencing increasing threats to their key roles in physical and ecological processes, including increased erosion from land use changes, sediment trapping by dam building, and the cascading hydrological and sedimentological effects of climate change (*1*). In many basins, the past and ongoing relative impacts of each of these changes remain unknown because of insufficient monitoring, and future projections are hindered by nonstationary watershed behavior, including changes without recent or well-monitored precedents (*2*, *3*). Given the divergent impacts of these different stressors—with land-use change dominantly increasing sediment flux, sediment trapping by dams dominantly decreasing sediment flux, and the effect of climate change on sediment flux uncertain—the temporal and global spatial variations in current and future net impacts are poorly constrained. We circumvented the global paucity of recent in situ river monitoring by using satellite analysis to investigate historic and ongoing changes in suspended sediment flux in 414 major rivers worldwide. We focused on sediment flux because of its essential role in maintaining riparian habitat and complexity, bolstering wetlands and deltas, delivering important nutrients to the oceans, and increasing coastal resilience to sea level rise. Because the major anthropogenic stresses to river functioning have both increased and decreased sediment flux, we found complex spatial and temporal variations in net impact on the world's major rivers, with major, global-scale reconfiguration in the past several decades.

[1]Department of Earth Sciences, Dartmouth College, Hanover, NH 03755, USA. [2]Department of Environmental Studies, Dartmouth College, Hanover, NH 03755, USA. [3]Department of Geography, Dartmouth College, Hanover, NH 03755, USA.
*Corresponding author. Email: evan.nylen.dethier@dartmouth.edu

We focused on 414 major rivers that flow into the oceans or major inland seas, are wider than ~90 m at their outlet, and have a drainage area of >20,000 km^2. Many of these rivers lack suspended sediment observations during the period of interest, and some have no published suspended sediment data at all. Recent advances in computational resources (*4*) and digital archives of hydrologic data allow for improved analysis of satellite image archives to monitor global water quality (*5*, *6*). We capitalized on these advances to evaluate decadal changes to river suspended sediment concentration (SSC) and flux, using a suite of algorithms trained on more than 130,000 ground truth measurements from 340 diverse sites in North America, South America, and Taiwan (*5*). Applied to the entire Landsat 5 and 7 archive, these algorithms allow for estimates of SSC for each purely water river pixel on Earth for hundreds to thousands of days from 1984 to present (fig. S1). In constructing a monthly record of SSC for each major river over time, we have filled in many of the critical gaps in the global understanding of spatial and temporal variability in river suspended sediment transport. For each river outlet, we extended our analysis with unified time series of SSC and river discharge estimates (*7*) to quantify global patterns and changes in discharge-weighted average SSC and monthly sediment flux to the oceans from 1984 to the present (Fig. 1) (*5*).

Dam building dominant in 20th century

We found that global river sediment flux is rapidly changing. Temporal trends in suspended sediment flux over the four-decade satellite record were significant for 53% of rivers (220 of 414; $P < 0.05$) (Fig. 2). Most rivers have retained at least a semblance of their natural variability and seasonality in SSC as, for example, SSC increases with increasing discharge (fig. S4), and the timing of peak SSC remains regionally consistent (Fig. 1 and fig. S4). However, our findings demonstrate in spatial and temporal detail the effects of direct human activity in altering suspended sediment transport, resulting in major changes to sediment transport in rivers and delivery of physical materials and nutrients to global estuaries and oceans. Decreasing sediment flux has been the prevailing global signal since the mid-20th century (Fig. 2), primarily because of dams (figs. S4 and S5), which are likely the dominant agent of riverine suspended sediment alteration in the past several centuries (*8*, *9*). However, this global signal masks the strong spatial variation in impacts. Widespread dam building in the 20th and 21st centuries has led to systematic declines in suspended sediment flux from rivers in the global hydrologic north, defined here geographically as north of ~20°N. Yet dam building is, at present, less pervasive in the global hydrologic south, defined here as south of ~20°N. There, intensive land-use change is driving rapidly increasing sediment fluxes. This strong north-south divergence is drastically reshaping the pattern of sediment delivery to global oceans (Fig. 3). Until the 1990s, Asia was the leading continental exporter of suspended sediment to the oceans and, during the pre-dam peak, accounted for ~50% of all sediment exported to global oceans; now, rivers in Asia export only 27 ± 1.9% of the global total.

The major Northern Hemisphere shift toward sediment retention in dammed reservoirs occurred in the 20th century. Of the rivers in this study, 58% (240 of 414) have one or more major dams (Fig. 4) (*10*), and 78% of those rivers (187 of 240) have decreased sediment flux relative to pre-dam estimates largely made by Milliman and Farnsworth (*11*). These pre-dam data are from a range of years and reflect varying estimation and extrapolation approaches and sampling techniques and periods. However, although these metadata gaps can make constraint of uncertainty challenging, the estimates are by far the most comprehensive pre-dam inventory of suspended sediment flux made from direct measurements and are an essential point of comparison. In aggregate, we found that continental fluxes are reduced by 49 ± 25%. The rivers with major dams are often comprehensively dammed. On average, reservoirs on dammed rivers have the capacity to retain 102% of annual river discharge (we refer to this normalized reservoir capacity simply as "reservoir capacity"). Continental sediment flux is decreased by 1.8 ± 0.6% per percent increase in reservoir capacity ($P = 0.043$) (Fig. 4), a negative trend that is also present and significant for individual rivers ($P = 0.002$) (fig. S10).

Dams have continued to reduce suspended sediment fluxes since 1984, although this effect has been strongly regional, focused where

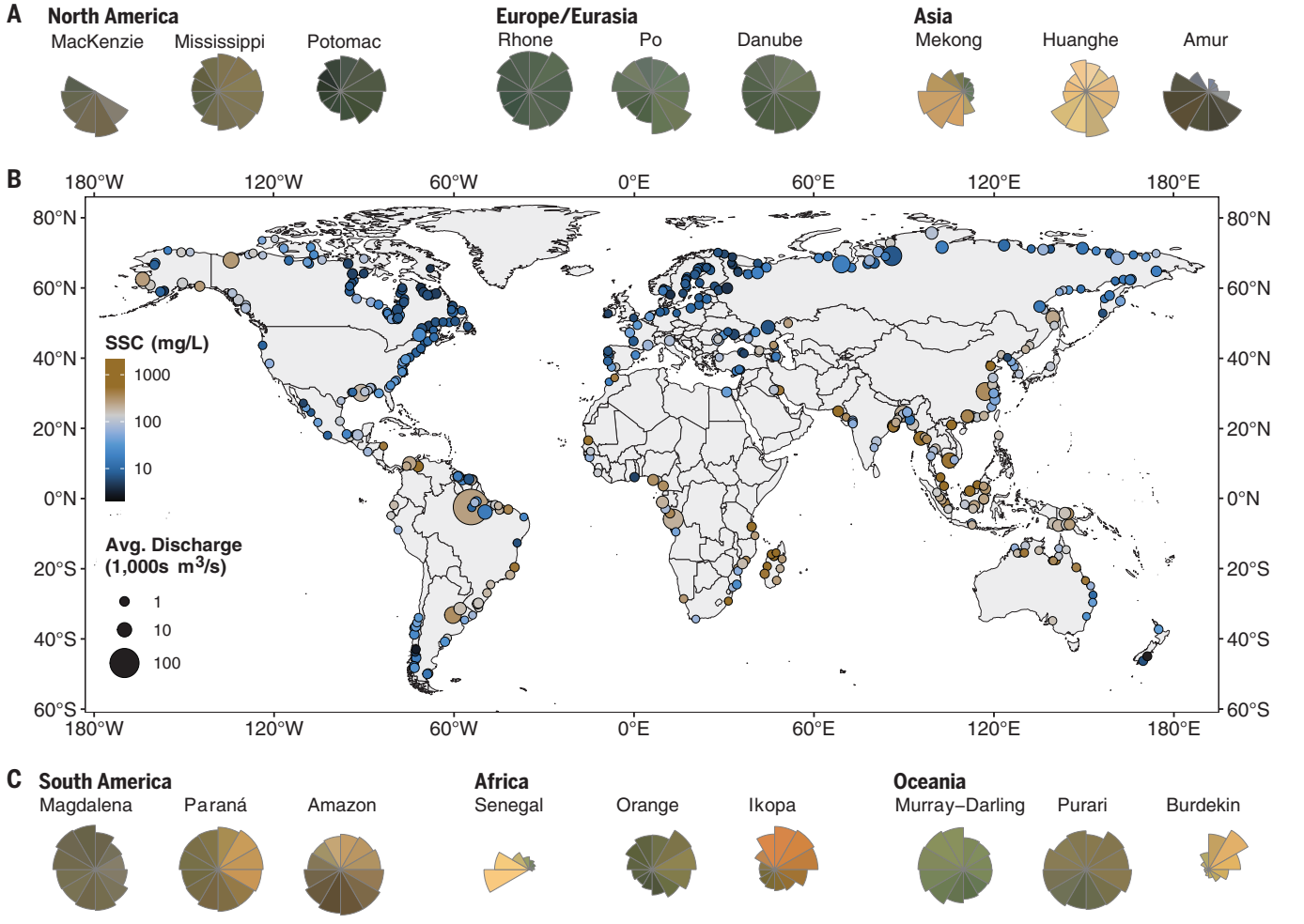


Fig. 1. Global variations in SSC, as estimated from satellite imagery from 1984 to 2021. (**A** and **C**) Variation in suspended sediment is shown with color wheels for select rivers from each primary landmass. Each wheel displays the monthly patterns in average river true color in Landsat imagery, by using the average red-green-blue (RGB) reflectance at the river outlet (supplementary materials, materials and methods). 1 January is at the top of each wheel, and each section of the wheel is scaled by monthly average discharge normalized by average peak monthly discharge for that river. Missing sections indicate ice-in season for high-latitude rivers. (**B**) Global map of average discharge-weighted SSC for 414 major rivers shows global patterns in SSC. Symbols are scaled by average annual discharge.

continued dam construction has occurred in Eurasia and Asia. Since 1984, approximately two-thirds (68%) of rivers with significant trends have decreased significantly (36% of all rivers), compared with increases at approximately one-third (32%) of rivers with significant trends (17% of all rivers). Reductions since 1984 have primarily occurred in Asia (−34 ± 7%) and Europe/Eurasia (−19 ± 12%) (Figs. 2A and 4B), where dam building on important undammed river reaches and tributaries continued through the mid-2000s (Fig. 4A).

However, in many ways our findings illustrate the diminishing additional impact of new dams on sediment flux to the global oceans. New dams and changing land-use practices continue to reduce sediment export from rivers in the global hydrologic north, primarily in Eurasia and Asia (Fig. 2, A and B, and, for example, fig. S8). However, fluxes from these rivers have on average only declined 7.5 ± 3.5% since the 1980s (Fig. 3C, top), a fraction of the 60 to 80% decline we documented relative to their pre-dam condition (Fig. 4B). Of the 146 rivers with post-1984 dams, about half (52%) have no statistically significant decline in suspended sediment flux, which possibly reflects increases in erosion from these watersheds but likely is primarily because previously built dams in those watersheds were already trapping substantial amounts of sediment. Indeed, rivers with post-1984 dams were widely dammed before 1984. On average, pre-1984 reservoir capacity for rivers with dams was 88 ± 8.2% of annual discharge, compared with 102 ± 10% today. With sediment retention already so high before the period of satellite observation, potential sediment flux reductions in the satellite period were comparatively limited. As a result, the post-1984 reductions were largely continuations of reductions owing to the initial wave of dam building in the mid-20th century. Of the rivers with post-1984 dams, only 7.5% (11 of 146) were built on rivers without at least one preexisting major dam.

Recent sediment flux increases in the global hydrologic south

Contrary to the northern trend of decreasing sediment flux, sediment fluxes from rivers in the global hydrologic south are rapidly and systematically increasing because of intensive land-use changes, particularly in watersheds within 20° of the Equator. About one-third (36%) of the 146 rivers in the global hydrologic south are transporting significantly more sediment than in 1984 (Fig. 3B). This rate of increase is significantly higher than for rivers in the global hydrologic north, where only 7%

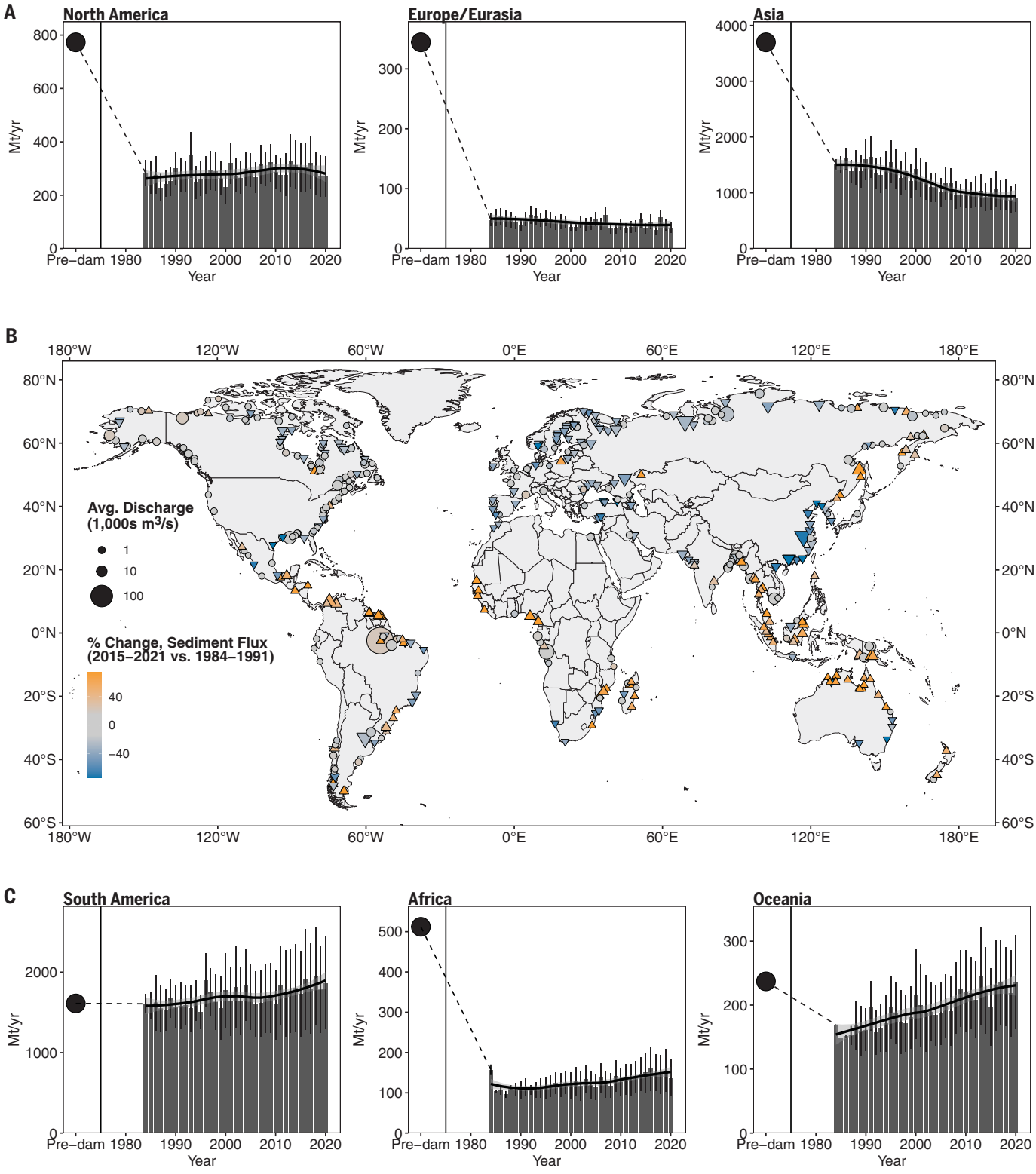


Fig. 2. Changes in suspended sediment flux for 414 major rivers on the six major global landmasses. (**A**) Changes for northern rivers relative to pre-dam fluxes estimated for those rivers (*11*). Annual estimates of suspended sediment flux show decrease or no change north of 20°N latitude, which is likely the result of continued dam building. (**B**) Rivers with increasing suspended sediment flux (upward-pointing triangles), rivers with declining suspended sediment flux (downward-pointing triangles), and rivers with no statistically significant change (circles). (**C**) Suspended sediment flux has increased south of 20°N over the same period owing to ongoing land-use change and intensification of the hydrological cycle. Scales for the *y* axes in (A) and (C) vary, and this analysis does not include every river on each landmass.

(18 of 268 rivers) have increasing sediment flux, despite increasing precipitation in several regions (fig. S7). Increases show heterogeneity in timing and magnitude, but discharge-weighted SSC has on average increased in southern rivers by 41 ± 7% (Fig. 3C, bottom). We attribute this increase in suspended sediment runoff from the global hydrologic south primarily to direct human influence. Suspended sediment flux increased after major land-use change in 58 of 70 total rivers with significant increasing trends (table S1), with most activities associated with widespread and/or acute deforestation in those watersheds (figs. S11 and S12) (*12*). In many cases, these land-use changes result in pronounced, point-source sediment inputs. Of particular note are alluvial mining, sand mining, and palm oil plantations (for example, fig. S12), which respectively appear to be primary drivers of increased erosion in 34, 19, and 17% of watersheds with increasing suspended sediment flux. Although sand mining is generally thought to reduce downstream bedload sediment flux, channel disruption and bank instability resulting from this practice could lead to increased suspended sediment load (*13*). These landscape-altering activities have become pervasive in tropical and subtropical watersheds in the past 20 years, often in or adjacent to rivers, without widespread recognition or documentation of their effects on rivers (fig. S5) (*1*, *13*–*15*).

We are unable to detect whether climate change and attendant changes to the hydrologic cycle have demonstrably contributed to changes in sediment supply. Some watersheds have experienced increased precipitation (*16*) or temperature-related erosion (*3*) during this period (fig. S6). However, although we observed

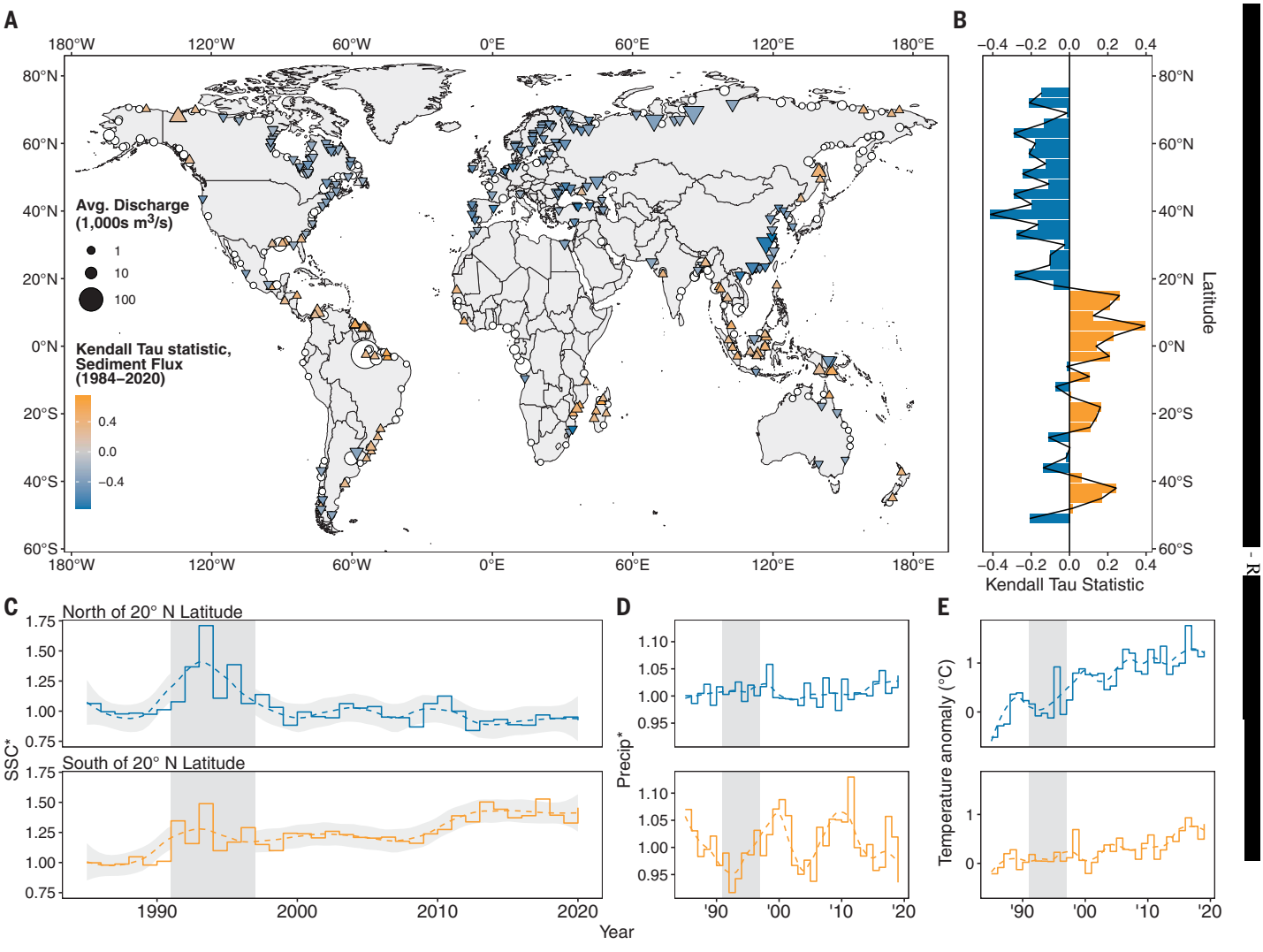


Fig. 3. Changes in suspended sediment flux from 1984 to 2020 at 414 major rivers. (**A**) Changes in river suspended sediment flux for each river, as determined with the Mann-Kendall nonparametric trend test. (**B**) Latitudinal variation in sediment flux trends, with the Mann-Kendall Tau statistic averaged for every river within 3° latitude bins, show a global divergence, with major increases south of 20°N and decreases north of 20°N. (**C**) Changes in SSC normalized by the 1984–1990 period (SSC*), with the Pinatubo eruption period removed, show a 7.5 ± 3.5% decline for northern rivers and a 41 ± 7% increase for southern rivers, which are likely driven by accelerating intensive land-use changes. (**D**) Changes in precipitation normalized by the 1984–1990 period (Precip*) do not show a significant relationship with trends in SSC*. (**E**) Changes in temperature relative to the 1984–1990 baseline show faster warming in northern watersheds, countering the hypothesis that increasing SSC for southern watersheds is the result of increased temperature. The gray bands in (C) and (D) indicate the standard error in the estimate for each year. Gray shading in (C) to (E) indicates years when the eruption of Mt. Pinatubo affected regional climates and satellite measurements of some rivers, particularly the upward excursion in SSC* for northern rivers.

cyclical changes in precipitation in southern-latitude watersheds, we found no correlation with the systematic change in SSC (P = 0.80) (Fig. 3D). Temperature, another potential explanatory variable because of its connection to rates of weathering and soil erosion (fig. S13) (*17*), has warmed more slowly in southern watersheds (~0.5°C) than in northern watersheds (~1.6°C) (Fig. 3E), where we observed stable or declining SSC. These potential or incipient effects of climate change on rivers do not yet pose a threat comparable with that of direct anthropogenic impacts. Yet they have been the major focus of international attention, including in the recent Intergovernmental Panel on Climate Change (IPCC) reports (*18*), possibly in part because of the lack of monitoring on tropical and subtropical rivers that account for many of the pronounced recent changes. The immediate consequences of direct land-use change, which are far greater in the present and near future, thus risk being overlooked.

Continued threats to rivers in the 21st century

In addition to being hotspots for intensive land-use change, for now the regions with increasing sediment fluxes remain notable exceptions to the global proliferation of dams and consequent decline in suspended sediment

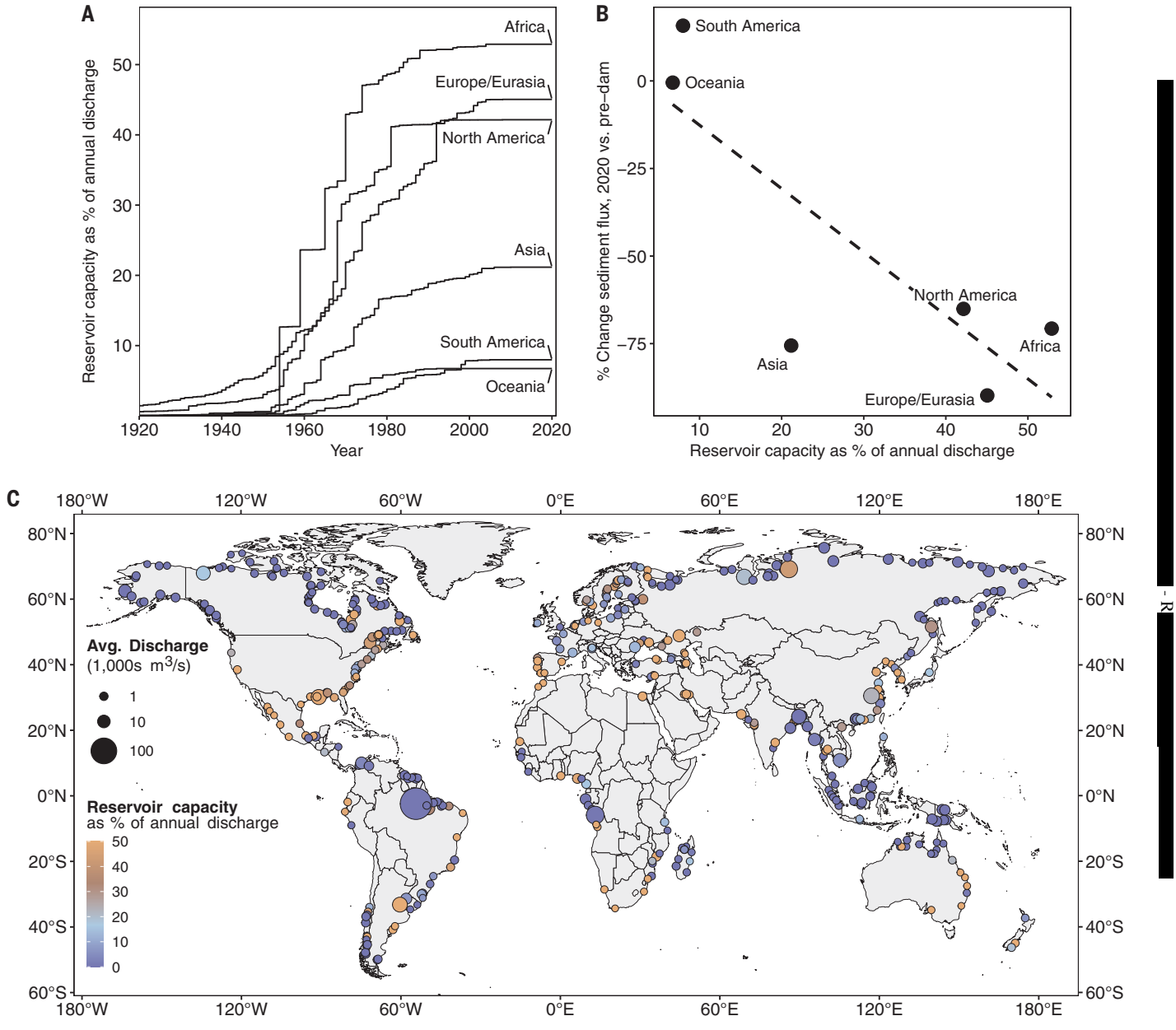


Fig. 4. Increases in reservoir capacity as a percent of annual discharge (RCI). (**A**) Dam building in Africa, North America, and Europe/Eurasia have capacity for ~40 to 50% of annual river flow, which is far greater than capacity in Asia (~20%), South America (~8%), and Oceania (~6%). (**B**) RCI is significantly negatively correlated with the change in suspended sediment flux relative to pre-dam (P = 0.043). Departure below the best-fit line in Asia is likely partially the consequence of a national soil conservation project and recent major dams missing from global dam inventories. (**C**) Map of RCI indicates the remaining undammed or minimally dammed rivers in dark blue, which are concentrated in the arctic and near the equator.

flux. South America and Oceania have the lowest reservoir capacity of all the continents. Reservoirs on those landmasses are capable of containing <10% of total river flow (Fig. 4), which is only a fraction of the reservoir capacity on other continents. Before global dam building, South America transported ~20% of the sediment from rivers in our study. Sediment fluxes from this continent have increased, in both relative and absolute terms, as declines have occurred in the Northern Hemisphere. South American rivers represent 52 ± 2.6% of the global total and, despite increases in river headwater sediment fluxes in Asia (*3*), are now the leading exporters of sediment to the oceans.

Because of ongoing and imminent dam building, the stable or increasing sediment fluxes to the ocean from South America and Oceania may not persist. To this point, dams have not wholly insulated river outlets from these changes; 40% of the southern rivers with increasing suspended sediment flux have at least one major dam, including 13% with dams built since 1984. However, more than 300 large dams are planned for the Amazon River, and more are planned for other rivers in South America and Oceania that are similarly unregulated (*19*, *20*). The Amazon, which exports two-thirds of the sediment from South America and more sediment than any other global river, is home to globally unrivaled channel, floodplain, and estuary biodiversity (*21*) and transfers essential nutrients to coastal waters (*22*). The planned dam projects there and on other unregulated rivers are particularly impactful because the current absence of dams allows virtually unchecked river transport of sediment, conditions that persist along few major rivers.

Elsewhere, dams are increasingly overprinted on systems that already face acute anthropogenic pressures, extending threats to critical human use and ecosystem services (*1*). Particularly in tropical and subtropical regions, losses of riverine function have accelerated because of morphological alteration (*13*, *15*) and physical and contaminant pollution (*15*, *23*, *24*). Planned dam projects on major rivers with existing dams—for example, more than 100 dams on the Mekong alone (*25*, *26*), as well as many on the Ganges/Brahmaputra, Yangtze, Danube, La Plata, and Uruguay Rivers, and dozens on other systems (*27*)—may have diminishing sediment trapping effects because those systems already have numerous major dams. In those cases, although successive dams might have less impact on global sediment export to the oceans, the upper- and middle-river channels and floodplains of these systems will nonetheless experience continued sediment reduction. Additionally, although sand mining has the potential to liberate sediment from channel beds, banks, and floodplains, the comprehensive removal of these materials for industrial uses may ultimately contribute to reduced suspended sediment loads (*13*, *28*). In reaches not affected by intensive land-use change, diminished sediment fluxes may alter channel migration rates (*29*), critical floodplain exchange (*30*), channel morphology (*31*), and in-channel habitat and biodiversity (*32*). Sediment retention in dam reservoirs can contribute to increases in downstream erosion (*33*), with potential consequences for downstream communities and estuaries.

Summary

We offer both a retrospective and current assessment of the global state of large rivers. We highlight the various pressing threats to these vital systems, especially the southward shift in the global center of mass of suspended sediment export to the oceans. The preponderance of both dams and rapid land-use change along the world's rivers continues their fundamental alteration through direct human action, so far outpacing the growing threats from climate change. As nations plan more dam projects, alter land-use practice, or reexamine current infrastructure and policy, stakeholders involved in planning decisions in the riparian corridor or coastal zone—for example, dam emplacement and removal, wetland resiliency, land-use management, and climate change adaptation—have historically made decisions with inadequate data (*19*). We have demonstrated the value of our satellite remote-sensing approach for estimating SSC and suspended sediment flux. Validated with 130,000 ground truth measurements, this technique can be updated and refined as additional in situ measurements are added to the calibration datasets and new satellites improve our monitoring coverage of Earth's surface (*34*). The real-time assessment of sediment transport by rivers can help inform policy decisions through direct observation of extant and historical conditions. Critically, we also provide the tools to evaluate the success of policy actions and interventions intended to improve conservation or public safety efforts. Our approach, as presented here and in the hands of other researchers and stakeholders, may help provide a robust means of impact assessment and assist in future planning, early detection, and mitigation efforts.

REFERENCES AND NOTES

1. J. Best, *Nat. Geosci.* **12**, 7–21 (2019).
2. P. C. D. Milly *et al.*, *Science* **319**, 573–574 (2008).
3. D. Li *et al.*, *Science* **374**, 599–603 (2021).
4. N. Gorelick *et al.*, *Remote Sens. Environ.* **202**, 18–27 (2017).
5. E. N. Dethier, C. E. Renshaw, F. J. Magilligan, *J. Geophys. Res.* **125**, e2019 (2020).
6. J. R. Gerson *et al.*, *Sci. Adv.* **6**, eabd4953 (2020).
7. B. M. Fekete, C. J. Vörösmarty, W. Grabs, *Global Biogeochem. Cycles* **16**, 15-1–15-10 (2002).
8. W. L. Graf, *Water Resour. Res.* **35**, 1305–1311 (1999).
9. H. Gupta, S. J. Kao, M. H. Dai, *J. Hydrol. (Amst.)* **464–465**, 447–458 (2012).
10. B. Lehner *et al.*, *Front. Ecol. Environ.* **9**, 494–502 (2011).
11. J. D. Milliman, K. L. Farnsworth, *River Discharge to the Coastal Ocean: A Global Synthesis* (Cambridge Univ. Press, 2013).
12. M. C. Hansen *et al.*, *Science* **342**, 850–853 (2013).
13. C. R. Hackney *et al.*, *Nat. Sustain.* **3**, 217–225 (2020).
14. M. Bendixen, J. Best, C. Hackney, L. L. Iversen, *Nature* **571**, 29–31 (2019).
15. E. N. Dethier, S. L. Sartain, D. A. Lutz, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23936–23941 (2019).
16. J. Barichivich *et al.*, *Sci. Adv.* **4**, eaat8785 (2018).
17. M. A. Nearing, F. F. Pruski, M. R. O'Neal, *J. Soil Water Conserv.* **59**, 43–50 (2004).
18. H.-O. P. Mbow, A. Reisinger, J. Canadell, P. O'Brien, "Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems (SR2)" (IPCC, 2017).
19. E. M. Latrubesse *et al.*, *Nature* **546**, 363–369 (2017).
20. K. O. Winemiller *et al.*, *Science* **351**, 128–129 (2016).
21. J. Salo *et al.*, *Nature* **322**, 254–258 (1986).
22. C. A. Nittrouer, D. J. DeMaster, *Cont. Shelf Res.* **16**, 553–573 (1996).
23. J. H. Mol, P. E. Ouboter, *Conserv. Biol.* **18**, 201–214 (2004).
24. K. Telmer, M. Costa, R. Simões Angélica, E. S. Araujo, Y. Maurice, *J. Environ. Manage.* **81**, 101–113 (2006).
25. G. M. Kondolf, Z. K. Rubin, J. T. Minear, *Water Resour. Res.* **50**, 5158–5169 (2014).
26. Z. Xue, J. P. Liu, Q. A. Ge, *Earth Surf. Process. Landf.* **36**, 296–308 (2011).
27. C. Zarfl, A. E. Lumsdon, J. Berlekamp, L. Tydecks, K. Tockner, *Aquat. Sci.* **77**, 161–170 (2015).
28. X. X. Lu, S. R. Zhang, S. P. Xie, P. K. Ma, *Hydrol. Earth Syst. Sci.* **11**, 1897–1906 (2007).
29. J. A. Constantine, T. Dunne, J. Ahmed, C. Legleiter, E. D. Lazarus, *Nat. Geosci.* **7**, 899–903 (2014).
30. T. Dunne, L. A. K. Mertes, R. H. Meade, J. E. Richey, B. R. Forsberg, *Geol. Soc. Am. Bull.* **110**, 450–467 (1998).
31. F. K. Ligon, W. E. Dietrich, W. J. Trush, *Bioscience* **45**, 183–192 (1995).
32. J. M. Friedman, W. R. Osterkamp, M. L. Scott, G. T. Auble, *Wetlands* **18**, 619–633 (1998).
33. G. M. Kondolf, *Environ. Manage.* **21**, 533–551 (1997).
34. S. Biancamaria, D. P. Lettenmaier, T. M. Pavelsky, *Surv. Geophys.* **37**, 307–337 (2016).
35. E. Dethier, evandethier/satellite-ssc: outlet-rivers. Zenodo (2022); doi: 10.5281/zenodo.6366616.

ACKNOWLEDGMENTS

We are grateful for the helpful contributions of T. Pavelsky and two anonymous reviewers. We thank H. Rubin, O. Lantz, G. Bolinger, and S. Stearns for their contributions to this project. **Funding:** This work was supported by the Dartmouth College Earth Science Department, Neukom Institute for Computational Studies, National Science Foundation EAR-1545623 (to E.N.D., C.E.R., and F.J.M.), National Science Foundation BCS-1636415 (to E.N.D., C.E.R., and F.J.M.), and National Aeronautics and Space Administration 80NSSC21K0309 (to E.N.D.). **Author contributions:** Conceptualization: E.N.D., C.E.R., and F.J.M. Methodology: E.D. Investigation: E.D. Visualization: E.D. Funding acquisition: C.R. and F.M. Project administration: E.N.D., C.E.R., and F.J.M. Writing – original draft: E.N.D., C.E.R., and F.J.M. Writing – review and editing: E.N.D., C.E.R., and F.J.M. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data used in this analysis are publicly available. Data and code used to generate all analysis, figures, and tables can be found at https://www.github.com/evandethier/satellite-ssc, with release archived at (*35*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn7980
Materials and Methods
Figs. S1 to S13
Table S1
References (*36*–*55*)

Submitted 20 December 2021; accepted 25 April 2022
10.1126/science.abn7980

MICROBIOLOGY

A centimeter-long bacterium with DNA contained in metabolically active, membrane-bound organelles

Jean-Marie Volland[1,2]*†, Silvina Gonzalez-Rizzo[3]†, Olivier Gros[3,4]*†, Tomáš Tyml[1,2], Natalia Ivanova[1], Frederik Schulz[1], Danielle Goudeau[1], Nathalie H. Elisabeth[5], Nandita Nath[1], Daniel Udwary[1], Rex R. Malmstrom[1], Chantal Guidi-Rontani[6], Susanne Bolte-Kluge[7], Karen M. Davies[5,8]‡, Maïtena R. Jean[3], Jean-Louis Mansot[4], Nigel J. Mouncey[1], Esther R. Angert[9], Tanja Woyke[1,2,10]*, Shailesh V. Date[2,11,12]*

Cells of most bacterial species are around 2 micrometers in length, with some of the largest specimens reaching 750 micrometers. Using fluorescence, x-ray, and electron microscopy in conjunction with genome sequencing, we characterized *Candidatus* (*Ca.*) Thiomargarita magnifica, a bacterium that has an average cell length greater than 9000 micrometers and is visible to the naked eye. These cells grow orders of magnitude over theoretical limits for bacterial cell size, display unprecedented polyploidy of more than half a million copies of a very large genome, and undergo a dimorphic life cycle with asymmetric segregation of chromosomes into daughter cells. These features, along with compartmentalization of genomic material and ribosomes in translationally active organelles bound by bioenergetic membranes, indicate gain of complexity in the *Thiomargarita* lineage and challenge traditional concepts of bacterial cells.

Bacteria and archaea are taxonomically and metabolically the most diverse and abundant organisms on Earth, but with only a small fraction of them isolated in culture, we remain grossly ignorant of their biology (*1*). Although most model bacteria and archaea are small, some remarkably large cells, referred to as giant bacteria, are evident in at least four phyla (*2*), and have cellular sizes in the range of tens or even hundreds of micrometers (*3*, *4*). Some exceptional members of the sulfur-oxidizing gammaproteobacteria *Thiomargarita namibiensis*, for instance, are known to reach up to 750 μm (average size: 180 μm) (*4*–*6*). Such bacterial giants raise the question of whether other lineages of previously unidentified macrobacteria might exist.

Here, we describe a sessile filamentous *Thiomargarita* species from a marine sulfidic environment that is larger than all other known giant bacteria by ~50-fold. Our multifaceted imaging analyses revealed massive polyploidy and a dimorphic developmental cycle in which genome copies are asymmetrically segregated into apparent dispersive daughter cells. We show that centimeter-long *Thiomargarita* filaments represent individual cells with genetic material and ribosomes compartmentalized into a metabolically active, membrane-bound organelle. Sequencing and analysis of genomes from five single cells revealed insights into distinct cell division and cell elongation mechanisms. These cellular features likely allow the organism to grow to an unusually large size and circumvent some of the biophysical and bioenergetic limitations on growth. In reference to its exceptional size, we propose to name this species *Thiomargarita magnifica*, which is hereafter referred to as *Candidatus* (*Ca.*) Thiomargarita magnifica.

Ca. T. magnifica is a centimeter-long, single bacterial cell

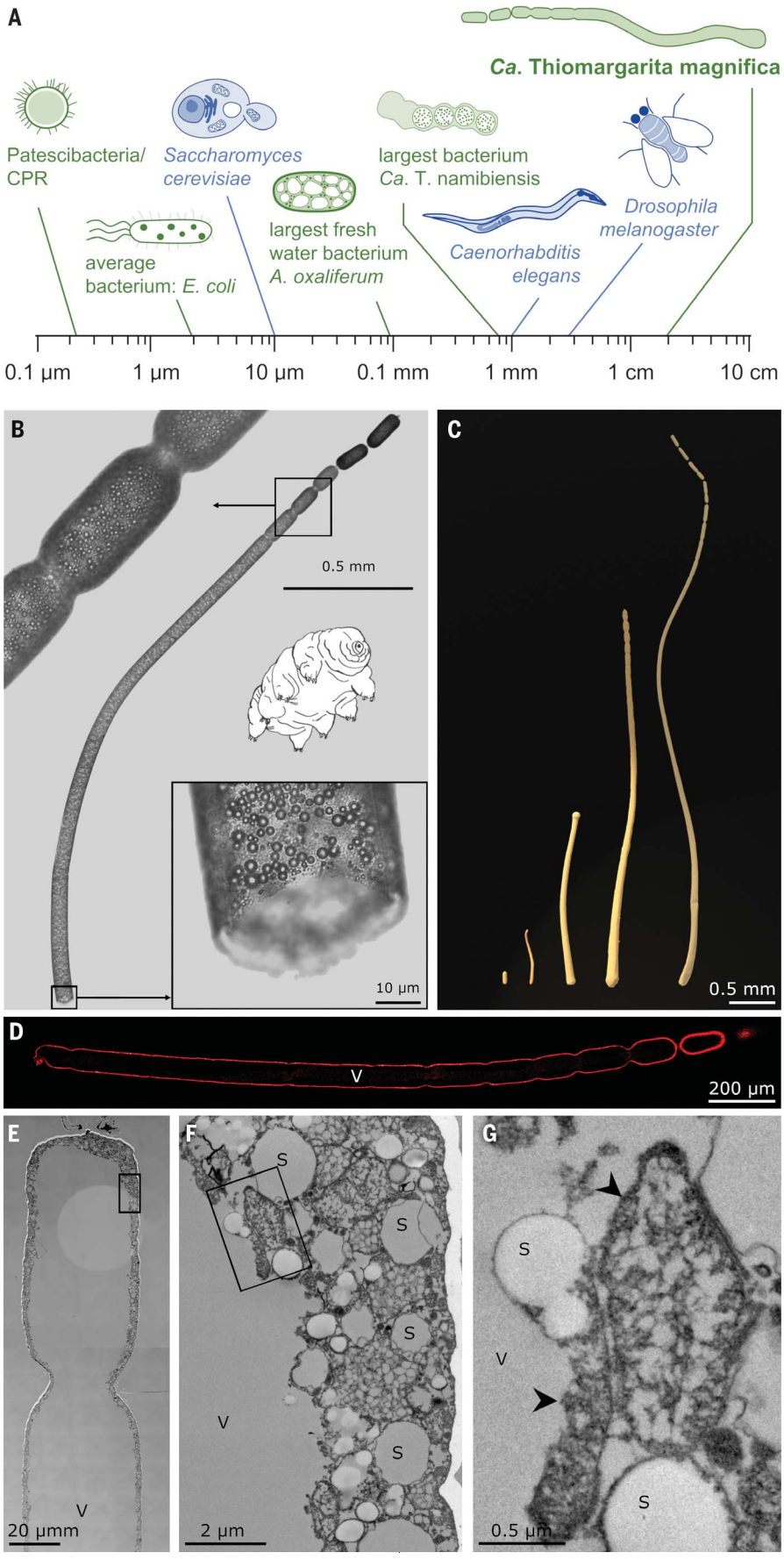
Some large sulfur bacteria (LSB) form very long filaments that may reach several centimeters in length but are composed of thousands of individual cells that do not exceed 200 μm (*7*–*10*). We observed seasonal "bouquets" of centimeter-long white filamentous *Thiomargarita* cells attached to sunken leaves of *Rhizophora mangle* (fig. S1) in shallow tropical marine mangroves from Guadeloupe, Lesser Antilles. *Thiomargarita* spp. are sulfur-oxidizing gammaproteobacteria known to be morphologically diverse and to display polyphenism (*11*). The morphology of the filaments observed in Guadeloupe resembled those of sessile *Thiomargarita*-like cells reported from deep-sea methane seeps (*12*). They had a stalk-like shape for most of their length and constricted gradually toward the apical end, forming buds (Fig. 1, A to E). In contrast to relatives that live buried in sediment, these filaments were smooth in appearance and free of epibiotic bacteria or any extracellular mucus matrix (figs. S1 and S2) (*11*). Budding filaments had an average length of 9.72 ± 4.25 mm, and only the most apical constrictions closed completely to form one to four separate, rod-shaped cells of 0.21 ± 0.05 mm. We also noted that some filaments reached a length of 20.00 mm (Fig. 1A and figs. S1 and S3), much larger than any previously described single-celled prokaryote.

To further characterize *Ca.* T. magnifica cells, we highlighted membranes using osmium tetroxide or the fluorescent dye FM 1-43X, and visualized entire filaments in three dimensions (3D) with hard x-ray tomography (n = 4; Fig. 1C and movies S1 and S2) and confocal laser scanning microscopy (CLSM) (n = 6; Fig. 1, C and D, and movie S3). Filament sections (up to 850.6 μm in length) were visualized with transmission electron microscopy (TEM) (n = 15; Fig. 1, E to G, and fig. S5). All techniques consistently showed that each filament was one continuous cell for nearly its entire length, with no division septa, including the partial constrictions toward the apical pole. Only the most apical few buds were separated from the filament by a closed constriction and these represented daughter cells (Fig. 1, fig. S5, and movies S1 to S4).

Similar to other LSB, *Ca.* T. magnifica cells has a large central vacuole that reduces the cytoplasmatic space. This may minimize growth limitation due to the reliance on chemical diffusion because bacteria lack an active intracellular transport system (*2*, *4*). In *Ca.* T. magnifica, the central vacuole was continuous along the whole filament and accounted for 73.2 ± 7.5% (n = 4) of the total volume (Fig. 1, D and E; fig. S5; and table S2). The cytoplasm was 3.34 ± 1.48 μm thick and was constrained to the periphery of the cell, so it was preserved from such chemical diffusion limitations (Fig. 1, E and F, and fig. S5) (*4*).

Within the cytoplasm, TEM revealed numerous lucent vesicles 2.40 ± 1.03 μm in diameter, which corresponded to the refractile granules observed with bright-field microscopy and represented sulfur granules, as shown by energy dispersive x-ray spectroscopy (Fig. 1F, figs. S5 and S13, and supplementary text). The cytoplasm of *Ca.* T. magnifica appeared to contain many electron-dense membrane-bound compartments 1.31 ± 0.70 μm in diameter (Fig. 1, F and G). Similar structures have

[1]Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [2]Laboratory for Research in Complex Systems, Menlo Park, CA, USA. [3]Institut de Systématique, Evolution, Biodiversité, Université des Antilles, Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Campus de Fouillole, Pointe-à-Pitre, France. [4]Centre Commun de Caractérisation des Matériaux des Antilles et de la Guyane, Université des Antilles, UFR des Sciences Exactes et Naturelles, Pointe-à-Pitre, Guadeloupe, France. [5]Department of Energy Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [6]Institut de Systématique, Evolution, Biodiversité CNRS UMR 7205, Museum National d'Histoire Naturelle, Paris, France. [7]Sorbonne Universités, UPMC Univ. Paris 06, CNRS FRE3631, Institut de Biologie Paris Seine, Paris, France. [8]Department of Molecular and Cell Biology, University of California, Berkeley, USA. [9]Cornell University, College of Agriculture and Life Sciences, Department of Microbiology, Ithaca, NY, USA. [10]University of California Merced, School of Natural Sciences, Merced, CA, USA. [11]University of California San Francisco, San Francisco, CA, USA. [12]San Francisco State University, San Francisco, CA, USA.
***Corresponding author. Email: jvolland@lbl.gov (J.-M.V.); Olivier.Gros@univ-antilles.fr (O.G.); twoyke@lbl.gov (T.W.); shailesh.date@lrc.systems (S.V.D.)**
†These authors contributed equally to this work.
‡Present address: Electron Bio-Imaging Centre, Diamond Light Source, Harwell Science and Innovation Campus, Didcot, UK.



occasionally been observed in other LSB and have been referred to as "blebs of cytoplasm," "putative endobionts," "intracytoplasmic structures which appear to contain nuclear material," or "membrane-enclosed cytoplasmic compartments with ribosomes and DNA fibrils," but their nature and function remained elusive in these primarily ultrastructural studies (*7*, *10*, *13*). We hypothesized that some of these membrane-bound compartments within the cytoplasm may contain dispersed genomic material because polyploidy is evident in many giant bacteria (*2*, *14*, *15*).

DNA and ribosomes within a membrane-bound bacterial organelle

Although bacteria were once presumed to be uncompartmentalized "bags of enzymes," recent studies have revealed the presence of organelles with functions as diverse as anaerobic ammonium oxidation, photosynthesis, and magnetic orientation (*16*). No bacteria or archaea are known to unambiguously segregate their genetic material in the manner of eukaryotes. There is, however, some evidence of membrane-bound nucleoids in one member of the Atribacteria, which has a compartment containing DNA that occupies most of the cell's volume (*17*). Plancotomycetes such as *Gemmata obscuriglobus* also have cytosolic

Fig. 1. Morphology and ultrastructure of *Ca.* T. magnifica. (**A**) Size comparison of selected bacterial (green) and eukaryotic (blue) model systems on a log scale. (**B**) Light microscopy montage of the upper half of a *Ca.* T. magnifica cell, with a broken basal part revealing a tube-like morphology due to the large central vacuole and numerous spherical intracellular sulfur granules (a tardigrade is shown for scale). (**C**) 3D rendering of segmented cells from hard x-ray tomography (movies S1, S2, and S6) and CLSM (movie S3) putatively at various stages of the developmental cycle. From left to right, 3D rendered cells *D*, *B*, *F*, *G*, and *D* (table S2). Note that the smallest stage corresponds to the cell *D* terminal segment and was added to the left for visualization purposes. (**D**) CLSM observation of cell *K* (table S2) after fluorescent labeling of membranes with FM 1-43X showing the continuity of the cell from the basal pole to the first complete constriction at the apical end. (**E**) TEM montage of the apical constriction of a cell, with the cytoplasm constrained to the periphery. (**F**) Higher magnification of the area marked in (E), with sulfur granules and pepins at various stages of development. (**G**) Higher magnification of the area marked in (F) showing two pepins (arrowheads). S, sulfur granule; V, vacuole.

compartments with DNA, and recent work has shown that these compartments represent deep invaginations of a Gram-negative cytoplasmic membrane rather than a closed, membrane-bound organelle (*16*).

4′,6-Diamidino-2-phenylindole (DAPI) staining revealed that the DNA in *Ca*. T. magnifica cells was concentrated in the membrane-bound compartments (Fig. 2, H to J, and fig. S6). Although we did not observe any connection of these compartments to the cell envelope, their mechanism of formation, which may include cell membrane invagination, is yet to be studied. They also harbored electron-dense structures 10 to 20 nm in size, similar to signatures of

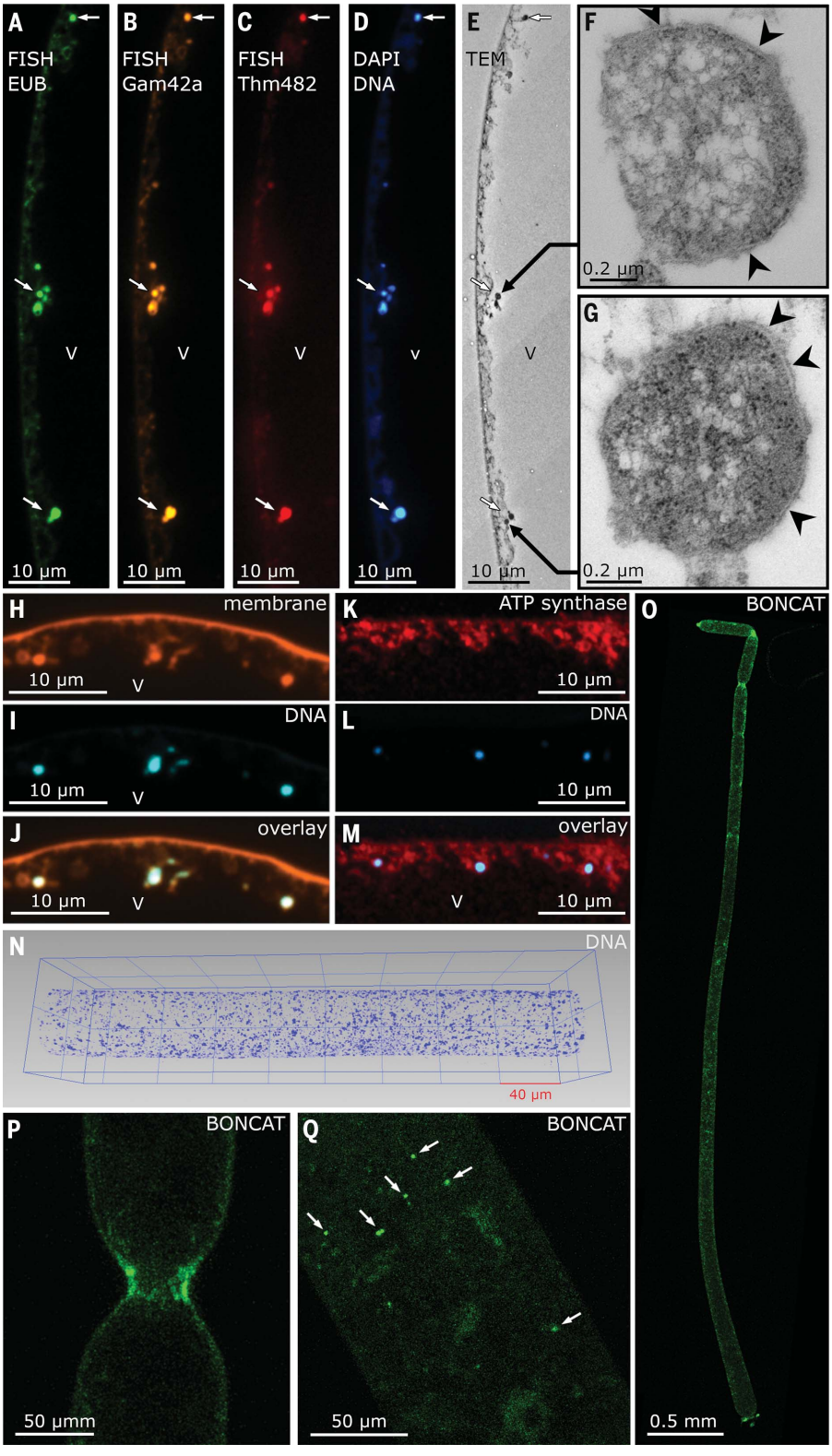


Fig. 2. Characterization of the pepin organelles by FISH and correlative TEM, as well as membrane and DNA staining, immunohistochemistry, and BONCAT. (**A** to **D**) FISH of pepins (arrows) in the cytoplasm of *Ca*. T. magnifica (class gammaproteobacteria). Pepins are labeled with the general bacterial probe EUB labeled with Alexa Fluor 488 [(A), green], the gammaproteobacteria-specific probe Gam42a labeled with Cy3 [(B), yellow], and the *Thiomargarita*-specific probe Thm482 labeled with Cy5 [(C), red] and with DAPI [(D), blue] (see the supplementary text for details). (**E**) TEM of a serial thin section consecutive to the semithin section used for FISH. The FISH- and DAPI-positive pepins appear as electron-dense organelles under TEM. (**F** and **G**) Pepins from (E) under higher magnification. Pepins are delimited by a membrane (arrowheads) and contain numerous ribosomes that appear as small, electron-dense granules throughout the sections of the pepins. (**H** to **J**) Fluorescent labeling of membranes using FM 1-43X (H) and of DNA using DAPI (I) and overlay (J) on a cross section of a cell. The pepins labeled with DAPI are also labeled with the dye FM 1-43X, confirming the presence of a membrane. (**K** to **M**) Visualization of ATP synthase localization using immunohistochemistry. Anti–ATP synthase antibodies label (K) reveals the presence of bioenergetic membranes around pepins [labeled with DAPI in (L)] and throughout the cytoplasm. (**N**) 3D visualization of a central portion of a cell after DAPI staining (blue) showing the multitude of DNA clusters spread throughout the cytoplasm (cell *M*; table S2 and movie S5). (**O** to **Q**) Translational activity revealed by BONCAT showing active protein biosynthesis within an entire cell, including hotspots at constriction sites and pepins [enlarged in (P) and (Q), respectively].

ribosomes (Fig. 2, F and G, and fig. S6, G to K). Fluorescence in situ hybridization (FISH) with probes specifically targeting ribosomal RNA sequences of *Thiomargarita* confirmed that ribosomes were indeed present and were concentrated in these membrane-bound structures (Fig. 2, A to G, and figs. S7 to S9) and spread throughout the entire cell, including the apical buds (fig. S9). This compartmentalization of DNA and ribosomes is reminiscent of the genomic nuclear compartmentalization in eukaryotes. By analogy with pips, the numerous small seeds in fruits such as watermelon or kiwi, we propose to name this bacterial organelle a *pepin* (singular *pepin*, plural *pepins*: from vulgar Latin *pép*, an expressive creation used to express smallness).

Cytoplasmic localization of bioenergetic membranes and translational activity

Adenosine triphosphate (ATP), the principal and universal energy currency of cells, is produced by ATP synthase, a molecular machine embedded in bioenergetic membranes and powered by proton motive force. Bacterial and archaeal ATP synthases are often observed to be localized to the cell envelope, in contrast to eukaryotes, in which ATP is generated by mitochondria. This cell envelope localization potentially constrains bacterial cell size because of the surface-to-volume ratio required to satisfy energy needs; a theoretical maximum cell size of ~10^{-14} m^3 has been estimated recently (*18*). Giant *Thiomargarita* cells are often cited as being exceptions to such energetic constraints and are thought to rely on their cell surface for ATP production (*19*–*21*). We used immunohistochemistry to assess the localization of ATP synthase in *Ca.* T. magnifica. We observed the distribution of ATP synthases around pepins and throughout the complex membrane network of the entire cytoplasm, but they were absent from the outer cell envelope (Fig. 2, K to M, and fig. S20).

To examine the localization of activity throughout the whole cell, we used bioorthogonal noncanonical amino acid tagging (BONCAT) to detect protein biosynthesis (*22*). Live filaments incubated with a clickable amino acid analog were labeled throughout (Fig. 2O and fig. S10). Consistent with the detection of ribosomes by FISH and TEM, BONCAT showed protein biosynthesis activity in small, round-shaped areas that were similar to pepins in size and localization, but not all pepins appeared to be labeled (Fig. 2Q and fig. S10D). Labeled hotspots were also observed at the site of constriction in the apical part of cells, suggesting higher translational activity or concentration of newly synthesized proteins in these areas (Fig. 2P). The establishment of stable continuous laboratory cultures of *Ca.* T. magnifica will likely be necessary to undertake detailed studies of the formation, biochemistry, and functions of pepins.

Based on these data, the metabolically active biovolume of *Ca.* T. magnifica (excluding the central vacuole) is two orders of magnitude above the predicted maximum mentioned above (table S2). This does not appear to contradict previous studies; indeed, such models have excluded bacteria with structural adaptations such as endomembrane systems and slower growth rates, which allow much larger cell volumes (*18*). Whereas most bacteria have doubling times ranging from minutes to hours, *Ca.* T. magnifica may be similar to other *Thiomargarita* species, which require up to 2 weeks to produce daughter cells (*5*). The predicted maximum volume of bacteria assumes binary fission as a division mode, but *Ca.* T. magnifica does not have to double its volume to produce a daughter cell because only a small portion of the apex constricts and detaches from the mother cell.

A highly polyploid cell with a large genome

All previously described giant bacteria are polyploid (*2*, *3*, *14*), meaning that their cells contain large numbers of genome copies ranging from tens to tens of thousands that are dispersed throughout the cell, supporting the local need for molecular machineries and overall cellular growth (*15*, *23*, *24*). Polyploidy has been shown to decrease the selective pressure on genes, allowing intracellular gene duplication, reassortment, and divergence, and to lead to extreme intracellular genetic diversity in some LSB (*25*). Conversely, polyploidy may allow balancing of genome copies through homologous recombination and support a high level of genome conservation (*26*). *Ca.* T. magnifica, like all bacterial giants, appeared to be polyploid; counts of DAPI-stained DNA clusters on three CLSM 3D datasets suggested an average of 36,880 ± 7956 genome copies per millimeter of filament (737,598 ± 159,115 for a fully grown 2-cm cell; see table S2, Fig. 2N, movie S5, and details in supplementary text). With its number of genome copies being one order of magnitude above that of other giant bacteria (*2*, *24*), *Ca.* T. magnifica accounts for the highest estimated number of genome copies within a single cell. Understanding the mechanisms underlying the regulation of such a large number of genome copies will require additional work.

To genomically characterize *Ca.* T. magnifica, we amplified, sequenced, and assembled the DNA of five individual cells collected from a single sunken leaf (tables S3 and S4). All five draft genomes were highly similar to each other, with an average nucleotide identity >99.5% (table S5). Although extreme intracellular genetic diversity has been shown in some LSB (*25*), our variant analysis of DNA sequences recovered from a single cell revealed a largely homogenous genome population (1.22 ± 0.18 single-nucleotide polymorphisms/100 kbp; table S6) (*27*), which is similar to other polyploid bacteria (*26*, *28*). The *Ca.* T. magnifica genome assemblies were estimated to be nearly complete at 91.0 to 93.7%, with total sequence lengths between 11.5 and 12.2 Mb. This value is twice as large as the only other sequenced *Thiomargarita* species, *Ca.* T. nelsonii (*29*, *30*), and at the upper range of bacterial genome sizes; bacterial genomes are on average 4.21 ± 1.77 Mb (fig. S11). The *Ca.* T. magnifica genomes contained up to 11,788 genes (only half with a functional annotation; table S4), more than three times the median gene count of prokaryotes (3935 genes) (*31*). For comparison with eukaryotic organisms, *Ca.* T. magnifica has a genome as large as the baker's yeast *Saccharomyces cerevisiae* (12.1 Mb) and contains more genes than the model fungus *Aspergillus nidulans* (≈ 9500 genes).

Analysis of the *Ca.* T. magnifica genome revealed a large set of genes for sulfur oxidation and carbon fixation, suggesting chemoautotrophy, which is consistent with other evidence for thioautotrophy (figs. S12 to S14 and supplementary text). Like its sister lineage *Ca.* T. nelsonii, *Ca.* T. magnifica encoded a wide range of metabolic capabilities with one notable difference: It lacked nearly all genes involved in dissimilatory and assimilatory nitrate reduction and denitrification except for Nar and Nap nitrate reductases (fig. S12). This absence suggests that nitrate can solely be used as an electron acceptor to be reduced to nitrite, which is not further reduced (*29*, *30*) (see the supplementary text for extended genome analysis). The absence of epibiotic bacteria may be explained by the high number of genes encoding secondary metabolism. With 25.9% of sequences dedicated to biosynthetic gene clusters (Fig. 3A), the genome encoded dozens of modular nonribosomal peptide synthetase and polyketide synthase systems, hinting at numerous secondary metabolism pathways (similar to the Actinobacteria; see table S7) that are indicative of antibiotic or bioactive compound production (*32*).

Giant bacteria have been shown to repurpose their cell division machinery as an adaptation to extreme cell size (*33*). The *Ca.* T. magnifica genome also holds clues for its unusual cell morphology in the form of an atypical complement of cell division and elongation genes. Many genes that encode core cell division proteins, including the core components of Z ring assembly and regulation, FtsA, ZipA, and FtsE-FtsX, were lacking (Fig. 3B and table S8). By contrast, genes that encode the cytoskeletal protein FtsZ, which is part of the well-conserved *dcw* ("division and cell wall") operon and the core component of the Z ring, were conserved. Proteins ZapA, ZapB, and ZapD, which interact with FtsZ and regulate Z ring assembly, were likewise conserved

(*34*).The entire set of genes that encode late divisome proteins, including peptidoglycan polymerases FtsI and FtsW, as well as FtsQ, FtsL, FtsB, and FtsK, was absent from all *Ca.* Thiomargarita genomes (Fig. 3B, fig. S19, and table S8). In contrast to the conspicuous lack of cell division genes, a complete set of genes encoding cell elongation proteins was present, three of which, *mreD*, *rodZ*, and the peptidoglycan transpeptidase *mrdA*, have undergone recent duplications, with both copies located next to each other on the chromosome (Fig. 3B, figs. S16 to S18, and table S8) (*34*). It is possible that an increased number of cell elongation genes, coupled with the lack of key cell division genes, may be responsible for producing the unusually long filaments of *Ca.* T. magnifica (see the supplementary text).

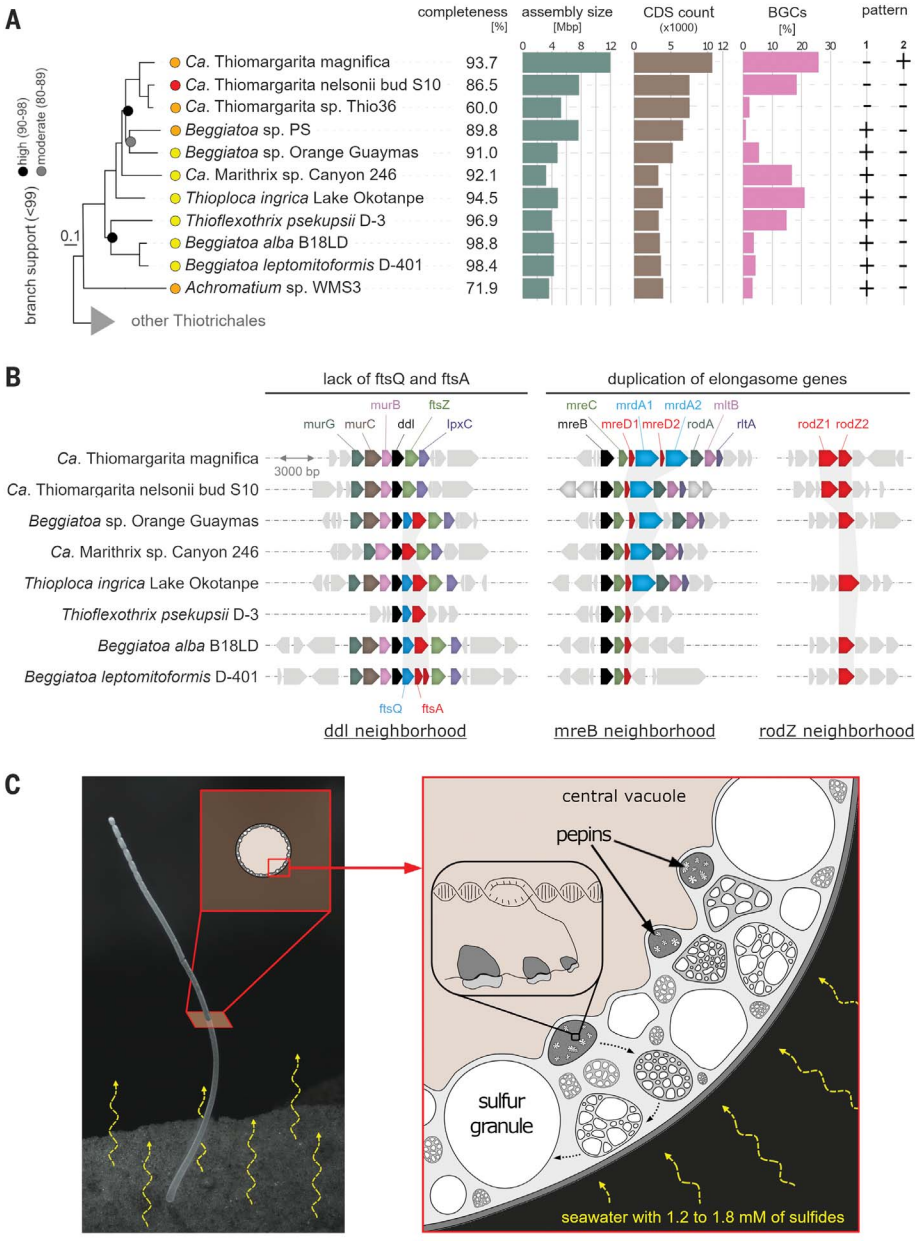


Fig. 3. Genome analysis and proposed model for the subcellular organization of *Ca.* T. magnifica. (**A**) Genome phylogenetic tree with added information about genome quality [red: low quality, orange: medium quality, and yellow: high quality (*46*)], estimated level of completeness, assembly size, coding sequence (CDS) count, and percentage of sequence dedicated to biosynthetic gene clusters (BGCs). Pattern 1 corresponds to "complete gene cluster for cell division of model bacteria." Pattern 2 corresponds to "*mreD*, *mrdA*, and *rodZ* genes are duplicated." (**B**) Gene neighborhoods centered on the *ddl*, *mreB*, and *rodZ* genes showing the incomplete set of divisome genes (lack of *ftsQ* and *ftsA*) in both *Thiomargarita* species, as well as the duplication of elongasome genes (*mreD*, *mrdA*, and *rodZ*) in *Ca.* T. magnifica. Note that the *Beggiatoa* sp. PS, *Achromatium* sp. WMS3, and *Ca.* Thiomargarita sp. Thio36 draft genomes were too fragmented and thus are not included here. (**C**) Light microscopy image and model proposed for the subcellular organization in *Ca.* T. magnifica showing how the pepin organelles might develop into other cellular compartments, resulting in an increase of surface area of the bioenergetic membranes.

Dimorphic developmental cycle of Ca. T. magnifica

Laboratory observations of live *Ca.* T. magnifica revealed eventual apical bud detachment from the filament and release into the environment, likely representing a dispersive stage of the developmental cycle (Fig. 1C; fig. S1, B to F; and supplementary text). We observed dozens of cells at all intermediary stages, from the smallest attached cells resembling terminal segments recently settled to the largest filaments with apical constrictions (fig. S1 and movies S1 to S3 and S6). Such a dimorphic life cycle resembles the aquatic single-celled model system *Caulobacter crescentus*, as well as the multicellular segmented filamentous bacteria, albeit at a different scale, in which stalked "parent" cells produce free-living "daughter" cells (*35*, *36*). Because of this asymmetrical division mode, only a small fraction of the genome copies present within pepins in the most apical bud (≈1%) were transmitted to the daughter cell (fig. S9). Like the polyploid giant bacterium *Epulopiscium* spp., *Ca.* T. magnifica apparently transmits only a subset of its genomes, so-called "germline genomes," to the offspring (*14*, *24*). If terminal buds are indeed daughter cells, then such a developmental cycle may have evolved to enhance dispersion similar to the fruiting bodies of the social myxobacteria or to the aerial hyphae of *Streptomyces* spp. (*37*). This apparent life cycle is also somewhat analogous to the sulfur-oxidizing giant ciliate symbiosis, *Zoothamnium niveum* (*38*), possibly representing a case of convergent evolution of developmental cycle across domains (see the supplementary text). As with other aspects of *Ca.* T. magnifica biology, detailed investigations of cell division and its regulation will require the establishment of stable laboratory cultures, and considering the cell size, spatial omics approaches within a single cell might be tractable.

Concluding remarks

Confirmation bias related to viral size prevented the discovery of giant viruses for more than a century, and their ubiquity is only now being recognized (*39*, *40*). The discovery of *Ca.* T. magnifica suggests that large and more complex bacteria may be hiding in plain sight.

Investigating the biology, energy metabolism, and the formation, nature, and role of pepins will take us a step closer to understanding the evolution of biological complexity.

Although cells of most bacteria and archaea are ~2 μm, eukaryotic cells are usually between 10 and 20 μm, with some of the largest single-cell eukaryotes reaching 3 to 4 cm (*41*). Several theoretical frameworks explain the restriction of bacteria and archaea to microscopic sizes, including: (i) the lack of active intracellular transport and the reliance on chemical diffusion, which is efficient only along micrometer distances (*4*); (ii) a predicted maximum cell volume constraining the number of needed ribosomes should the cell grow larger (*21*); or even (iii) a decrease in energy efficiency due to mismatched surface area to volume ratio when considering placement of membrane-bound ATP synthases (*18*, *20*). These frameworks all suggest that with increasing size, the physiological or metabolic needs of a bacterial cell grow faster than the cell's capacity to sustain it and should reach a limit. The next largest prokaryote known after *Ca.* T. magnifica, *Ca.* T. nelsonii, has a metabolically active biovolume of 1.05×10^{-14} m^3 (excluding the central vacuole), close to the predicted maximum due to ribosome limitations, 1.39×10^{-15} m^3, and to the bioenergetic membrane limitation of 10^{-14} m^3 (*18*, *21*). Our precise 3D measurements on a 4.27-mm *Ca.* T. magnifica cell revealed a cytoplasm biovolume several orders of magnitude above that limit (5.91×10^{-12} m^3; table S2). It is possible that changes in spatial organization of cellular components, such as DNA and ribosome compartmentalization and rearrangement of the bioenergetic membrane system, may allow *Ca.* T. magnifica to overcome many such limitations (Fig. 3C).

Distributed in at least 23 phyla are 19 known types of bacterial organelles, of which only seven are membrane bound (*16*, *42*). Cyanobacteria can form multicellular, centimeter-long filaments and are capable of cell differentiation (*43*). Planctomycetes have special energy transduction organelles called anammoxosomes and a compartmentalized cell, and some are even capable of phagocytosis (*16*, *44*). The social Myxobacteria have large genomes and a complex developmental cycle, and are capable of moving and feeding cooperatively in predatory groups (*45*). Through its gigantic cell size, its large genome, and its dimorphic life cycle, but most importantly through its compartmentalization of genetic material in membrane-bound pepins, *Ca.* T. magnifica adds to the list of bacteria that have evolved a high level of morphological complexity. Because it segregates its genetic material in membrane-bound organelles, *Ca.* T. magnifica challenges our concept of a bacterial cell.

REFERENCES AND NOTES

1. A. D. Steen *et al.*, *ISME J.* **13**, 3126–3130 (2019).
2. D. Ionescu, M. Bizic, "Giant bacteria," in *eLS* (Wiley, 2019); https://doi.org/10.1002/9780470015902.a0020371.pub2.
3. P. A. Levin, E. R. Angert, *Cold Spring Harb. Perspect. Biol.* **7**, a019216 (2015).
4. H. N. Schulz, B. B. Jorgensen, *Annu. Rev. Microbiol.* **55**, 105–137 (2001).
5. H. N. Schulz *et al.*, *Science* **284**, 493–495 (1999).
6. E. R. Angert, K. D. Clements, N. R. Pace, *Nature* **362**, 239–241 (1993).
7. D. C. Nelson, C. O. Wirsen, H. W. Jannasch, *Appl. Environ. Microbiol.* **55**, 2909–2917 (1989).
8. M. Huettel, S. Forster, S. Kloser, H. Fossing, *Appl. Environ. Microbiol.* **62**, 1863–1872 (1996).
9. M. R. Jean *et al.*, *PLOS ONE* **10**, e0117832 (2015).
10. J. M. Larkin, M. C. Henk, *Microsc. Res. Tech.* **33**, 23–31 (1996).
11. V. Salman, J. V. Bailey, A. Teske, *Antonie van Leeuwenhoek* **104**, 169–186 (2013).
12. J. V. Bailey *et al.*, *ISME J.* **5**, 1926–1935 (2011).
13. S. Maier, H. Volker, M. Beese, V. A. Gallardo, *Can. J. Microbiol.* **36**, 438–448 (1990).
14. E. R. Angert, *Annu. Rev. Microbiol.* **66**, 197–212 (2012).
15. S. Schorn *et al.*, *FEMS Microbiol. Ecol.* **96**, fiz200 (2020).
16. C. Greening, T. Lithgow, *Nat. Rev. Microbiol.* **18**, 677–689 (2020).
17. T. Katayama *et al.*, *Nat. Commun.* **11**, 6381 (2020).
18. P. E. Schavemaker, S. A. Muñoz-Gómez, The role of mitochondrial energetics in the origin and diversification of eukaryotes. bioRxiv 465364 [Preprint] (2021); https://doi.org/10.1101/2021.10.23.465364.
19. N. Lane, *Curr. Biol.* **30**, R471–R476 (2020).
20. N. Lane, W. Martin, *Nature* **467**, 929–934 (2010).
21. C. P. Kempes, L. Wang, J. P. Amend, J. Doyle, T. Hoehler, *ISME J.* **10**, 2145–2157 (2016).
22. R. Hatzenpichler *et al.*, *Environ. Microbiol.* **16**, 2568–2590 (2014).
23. J. E. Mendell, K. D. Clements, J. H. Choat, E. R. Angert, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6730–6734 (2008).
24. E. R. Angert, *Genome Biol. Evol.* **13**, evab037 (2021).
25. D. Ionescu, M. Bizic-Ionescu, N. De Maio, H. Cypionka, H. P. Grossart, *Nat. Commun.* **8**, 455 (2017).
26. V. Salman-Carvalho, E. Fadeev, S. B. Joye, A. Teske, *Front. Microbiol.* **7**, 1173 (2016).
27. B. J. Shapiro, *Curr. Opin. Microbiol.* **31**, 116–123 (2016).
28. T. Woyke *et al.*, *PLOS ONE* **5**, e10314 (2010).
29. B. E. Flood *et al.*, *Front. Microbiol.* **7**, 603 (2016).
30. M. Winkel *et al.*, *Front. Microbiol.* **7**, 964 (2016).
31. K. Chiyomaru, K. Takemoto, *R. Soc. Open Sci.* **7**, 191859 (2020).
32. L. Katz, R. H. Baltz, *J. Ind. Microbiol. Biotechnol.* **43**, 155–176 (2016).
33. D. A. Miller, G. Suen, K. D. Clements, E. R. Angert, *BMC Genomics* **13**, 265 (2012).
34. A. J. F. Egan, J. Errington, W. Vollmer, *Nat. Rev. Microbiol.* **18**, 446–460 (2020).
35. S. A. Woldemeskel, E. D. Goley, *Trends Microbiol.* **25**, 673–687 (2017).
36. E. A. Hutchison, D. A. Miller, E. R. Angert, *Microbiol. Spectr.* **2**, 2.5.22 (2014).
37. D. Claessen, D. E. Rozen, O. P. Kuipers, L. Søgaard-Andersen, G. P. van Wezel, *Nat. Rev. Microbiol.* **12**, 115–124 (2014).
38. M. Bright, S. Espada-Hinojosa, I. Lagkouvardos, J. M. Volland, *Front. Microbiol.* **5**, 145 (2014).
39. C. Abergel, M. Legendre, J. M. Claverie, *FEMS Microbiol. Rev.* **39**, 779–796 (2015).
40. F. Schulz *et al.*, *Nature* **578**, 432–436 (2020).
41. J. Dumais, K. Serikawa, D. F. Mandoli, *J. Plant Growth Regul.* **19**, 253–264 (2000).
42. C. A. Kerfeld, C. Aussignargues, J. Zarzycki, F. Cai, M. Sutter, *Nat. Rev. Microbiol.* **16**, 277–290 (2018).
43. E. Flores, A. Herrero, *Nat. Rev. Microbiol.* **8**, 39–50 (2010).
44. T. Shiratori, S. Suzuki, Y. Kakizawa, K. I. Ishida, *Nat. Commun.* **10**, 5529 (2019).
45. J. Muñoz-Dorado, F. J. Marcos-Torres, E. García-Bravo, A. Moraleda-Muñoz, J. Pérez, *Front. Microbiol.* **7**, 781 (2016).
46. R. M. Bowers *et al.*, *Nat. Biotechnol.* **35**, 725–731 (2017).

ACKNOWLEDGMENTS

We are thankful to the following centers where the electron microscopy analyses were performed: the Centre Commun de Caractérisation des Matériaux des Antilles et de la Guyane in Guadeloupe, F.W.I., which is supported by The European Regional Development Fund, the Regional Council of Guadeloupe, and the French Research Department; the Electron Microscope Lab (EML) of the University of California Berkeley; the Electron Microscopy Resource in Donner at LBNL, Berkeley; and the FEI Eindhoven Center. We are particularly grateful to D. Jorgens at EML for advice and assistance in electron microscopy sample preparation and data collection. The x-ray tomography data were acquired at the Stanford Nano Shared Facilities at Stanford University (Palo Alto, CA), and we are particularly grateful to A. Vailionis for his technical support during hard x-ray tomography scan acquisitions. The confocal microscopy observations were performed at the Advanced Microscopy Facility at LBNL (Berkeley, CA). Preliminary confocal microscopy observations were acquired at the IBPS Imaging Facility, which is supported by "Conseil Regional d'Ile-de-France." We thank S. Volland for his help with 3D rendering animations and H. Maughan for copyediting this manuscript. The work (proposal: 10.46936/10.25585/60001074) conducted by the US Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a Department of Energy (DOE) Office of Science User Facility, is supported by the Office of Science of the DOE operated under contract no. DE-AC02-05CH11231. The Nagoya permit TREL1820249A/50 (unique identifier ABSCH-IRCC-FR-246822-1) can be consulted publicly online at: https://absch.cbd.int/en/search. **Funding:** This work was supported by the John Templeton Foundation (grant 60973 to J.-M.V., S.V.D., and T.T.), the Gordon and Betty Moore Foundation (grant GBMF7617 to J.-M.V., S.V.D., and T.T.), DARPA (award no. HR001120036 to J.-M.V., S.V.D., and T.T.); the DOE Office of Science (contract no. DE-AC02-05CH11231 to T.W., F.S., T.T., J.-M.V., K.M.D., and N.H.E.), and Region Guadeloupe (F.W.I. grant to M.R.J.). **Author contributions:** Conceptualization: O.G., S.G.R., J.-M.V., T.W., S.V.D., F.S., R.R.M., N.I., T.T., N.H.E.; Data curation: N.I., F.S., J.-M.V., D.U., S.G.R.; Formal analysis: J.-M.V., S.G.R., F.S., N.I., D.U., O.G.; Funding acquisition: T.W., S.V.D., O.G., J.L.M.; Investigation: J.-M.V., S.G.R., O.G., T.T., F.S., D.U., D.G., N.N., N.I., C.G.R., S.B.K., N.H.E., M.R.J., J.L.M.; Methodology: O.G., T.T., J.-M.V., S.V.D., S.G.R., D.G., N.N., R.R.M., T.W., C.G.R., S.B.K.; Resources: K.M.D., T.W., R.R.M., O.G., J.L.M., N.H.E., N.J.M.; Supervision: O.G., T.W., S.V.D.; Visualization: O.G., S.G.R., J.-M.V., F.S., N.I., D.U.; Writing – original draft: J.-M.V.; Writing – review and editing: J.-M.V., S.G.R., O.G., T.W., S.V.D., E.A., N.I., R.R.M. **Competing interests:** S.V.D. serves as the CEO of Sample Exchange. The remaining authors declare no competing interests. **Materials and data availability:** All data needed to evaluate the conclusions in this study are present in the main manuscript or the supplementary materials. The complete assemblies as well as the extracted *Ca.* T. magnifica draft genomes have been deposited in IMG (https://img.jgi.doe.gov/). The raw reads have been deposited in SRA (https://www.ncbi.nlm.nih.gov/sra). Accession numbers are provided in tables S3 and S4.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abb3634
Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 to S11
References (*47*–*83*)
Movies S1 to S6
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 16 October 2021; resubmitted 26 March 2022
Accepted 9 May 2022
10.1126/science.abb3634

REPORT

AGING

Slow and negligible senescence among testudines challenges evolutionary theories of senescence

Rita da Silva[1,2,3]*†, Dalia A. Conde[1,2,3], Annette Baudisch[3,4], Fernando Colchero[2,3,5]*

Is senescence inevitable and universal for all living organisms, as evolutionary theories predict? Although evidence generally supports this hypothesis, it has been proposed that certain species, such as turtles and tortoises, may exhibit slow or even negligible senescence—i.e., avoiding the increasing risk of death from gradual deterioration with age. In an extensive comparative study of turtles and tortoises living in zoos and aquariums, we show that ~75% of 52 species exhibit slow or negligible senescence. For ~80% of species, aging rates are lower than those in modern humans. We find that body weight positively relates to adult life expectancy in both sexes, and sexual size dimorphism explains sex differences in longevity. Unlike humans and other species, we show that turtles and tortoises may reduce senescence in response to improvements in environmental conditions.

How much can aging be altered, slowed, or brought to a halt altogether? In the past century, we have witnessed unprecedented increases in human longevity (*1*). Yet, research on humans and nonhuman primates shows that these improvements have resulted from averting early deaths and age-independent sources of mortality, not from reducing the rate of aging (*2*, *3*). The rate of aging is a measure of the speed at which the risk of mortality increases with age. It is the direct result of senescence, a gradual deterioration of bodily functions that manifests as an increase in mortality risk with age after sexual maturity (*4*). Current evolutionary theories of senescence state that, among all organisms with a clear separation between somatic and germline cell lineages, senescence is inevitable (*4*, *5*). Paradoxically, empirical evidence (*6*, *7*) and evolutionary demographic models (*8*, *9*) have proposed that evolution may permit some species to reduce or even avoid the effects of senescence (i.e., negligible senescence).

Species that continue growing after reproductive maturity (e.g., turtles and tortoises) (*8*) are the prime candidates for escaping senescence. These indeterminately growing species may gain survival advantages and larger reproductive potential with age, which allows them to invest more in somatic maintenance and potentially slowing senescence. To date, only a handful of studies have investigated senescence in animal species with indeterminate growth, such as turtles and tortoises (*10*–*13*), where different populations of the same species can show evidence of both senescence and negligible senescence (*12*–*15*). Thus, the question remains: Can some species slow or even avoid growing old? And if so, under what circumstances?

In this work, we carried out an extensive study of age- and sex-specific mortality and growth patterns in turtles and tortoises (order Testudines). Using the Species360 Zoological Information Management System (ZIMS) (*16*), we obtained husbandry records for 52 species spanning a diversity of life-history strategies, body weights, and longevities (table S1). Using Bayesian survival trajectory analysis (*17*, *18*), we estimated for females (47 species) and males (39 species) adult age-specific mortality, remaining adult life expectancy, and aging rates. From the best-fitting models, we calculated 95% credible intervals (CIs) of aging rates at the age when the survival function reached 0.2 (i.e., when 80% of adults are expected to have died) (*19*). We considered this age to be sufficiently advanced to occur after the onset of senescence but not so late as to greatly increase the uncertainty in the estimated aging rates.

CIs of aging rates included zero for 74.5% of species (35 species) for females and 79.5% (31 species) for males (Fig. 1). CIs of some species were either negative [i.e., *Testudo graeca* and *Siebenrockiella crassicollis*, 4.2% (2 species) for females and 2.6% (1 species) for males] or spanned narrowly around zero (e.g., females of *Aldabrachelys gigantea* and males of *Gopherus berlandieri*), which may suggest the existence of negligible senescence among these species. CIs were positive for 21.3% (10 species) for females and 17.9% (7 species) for males, which suggests that these species experience senescence even under protected conditions. CIs were lower than the aging rate of modern humans for 78.7% of species (37 species) for females and 87.2% (34 species) for males (Fig. 1). The distribution of aging rates in testudines was considerably narrower than that for mammals, with minimal overlap between the two distributions (fig. S1). In short, aging rates among testudines fall much below those of most mammals and, for some species, are close to negligible (i.e., are not different from zero; fig. S2, A to D).

Using Bayesian phylogenetic generalized least-squares (PGLS) (*19*), we found that adult life expectancy was positively related to body weight for both sexes, as expected from life-history theory (*20*). Sexual size dimorphism, calculated as the difference in body weight between females and males relative to that of females, related positively to adult male life expectancy (i.e., males are expected to live longer for species where females are the larger sex), possibly owing to a lower risk of male-male aggression among species with female-biased sexual size dimorphism (*21*). With respect to aging rates, we found a marginal positive effect of sexual size dimorphism for both sexes but not of body weight. Because of a lack of accurate reproductive data at older ages, we were unable to test for a potential indirect effect of reproductive output on adult life expectancy and aging rates (*22*).

To further understand how longevity and aging rates vary between sexes, we measured the relative sex difference in adult life expectancy [$\delta_e = (e_{female} - e_{male}) / e_{female}$] and the difference in aging rate ($\delta_a = a_{female} - a_{male}$) for 34 species (Fig. 2) (*19*). Adult male life expectancy exceeded that of females by 20% (±51%) on average, whereas the average sex difference in aging rate was close to zero (−0.002 ± 0.039). CIs of adult life expectancy difference did not include zero for five species with male advantage and two with female advantage. Our finding of a male adult life expectancy advantage is congruent with studies of other reptilian taxa (*23*) but not with results on other tetrapods, such as mammals (*24*). In contrast to previous studies on reptiles (*21*, *25*–*27*), our Bayesian PGLS showed that sex differences in adult life expectancy could not be explained by differences in terrestrial versus aquatic habitats, differences in reproductive effort between sexes, or hibernation (Table 1). However, under human care, differences in habitat are negligible, and individuals may naturally bypass hibernation in response to unlimited access to food,

[1]Department of Biology, University of Southern Denmark, 5230 Odense M, Denmark. [2]Species360 Conservation Science Alliance, Bloomington, MN 55425, USA. [3]Interdisciplinary Centre on Population Dynamics, University of Southern Denmark, 5230 Odense M, Denmark. [4]Danish Centre for Population Research, University of Southern Denmark, 5230 Odense M, Denmark. [5]Department of Mathematics and Computer Science, University of Southern Denmark, 5230 Odense M, Denmark.
*Corresponding author. Email: colchero@imada.sdu.dk (F.C.); a.ritasilva.15@gmail.com (R.d.S.)
†Present address: CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado and BIOPOLIS Program in Genomics, Biodiversity and Land Planning, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal.

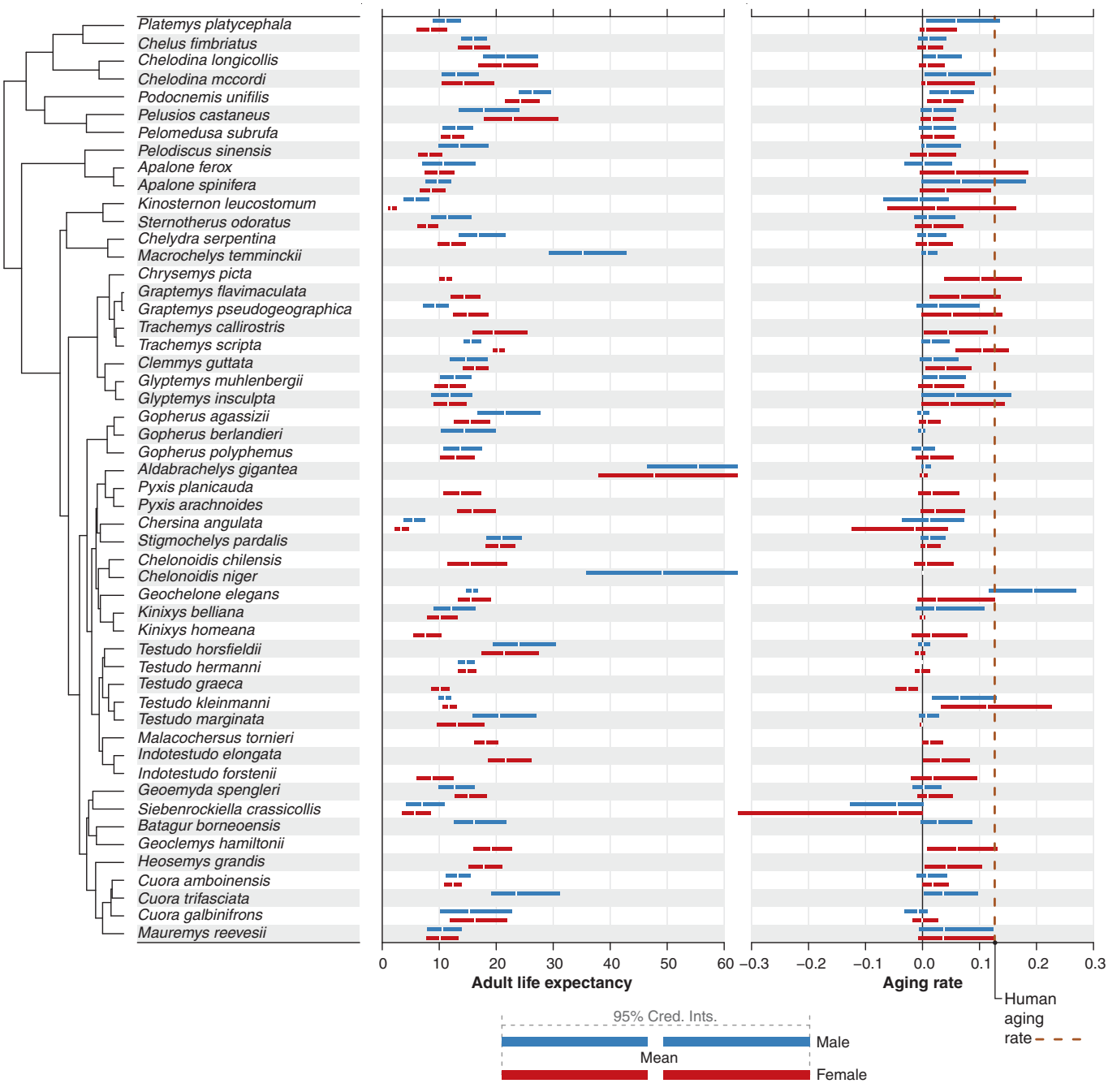


Fig. 1. Adult life expectancy and aging rates for female and male turtles and tortoises. Life expectancies and aging rates and their 95% CIs were calculated using Bayesian survival trajectory analysis (BaSTA). For comparison, we depict modern human aging rates. Cred. Ints., credible intervals.

protection from predation, and thermal stability year-round. Because of the lack of information, we could not test other explanations, such as the costs of sexually selected traits (*28*) or sex determination system (*29*).

Of the variables tested, only sexual size dimorphism strongly related to sex differences in adult life expectancy, whereby our model predicted that females had longer life expectancy when they were larger than males. A similar relationship among anurans has been attributed to sexual bimaturation (*30*). In such cases, females achieve a selective advantage by delaying maturation and attaining larger body sizes, slower growth, and longer life expectancies than males. We were unable to test this hypothesis because of the lack of information on sex-specific ages at maturity. Additionally, being larger than males could benefit females because courtship and mating can be physically harmful, particularly for male-biased sexual size dimorphism (*21*). Accentuated by confinement under human care, we hypothesize that sexual size dimorphism may affect life span in females through its influence on the longevity costs of courtship and mating. To test this hypothesis, future analyses should compare aging rates of populations where sexes are separated against those where they are kept together.

In this work, we have studied populations under protected environments, and thus our

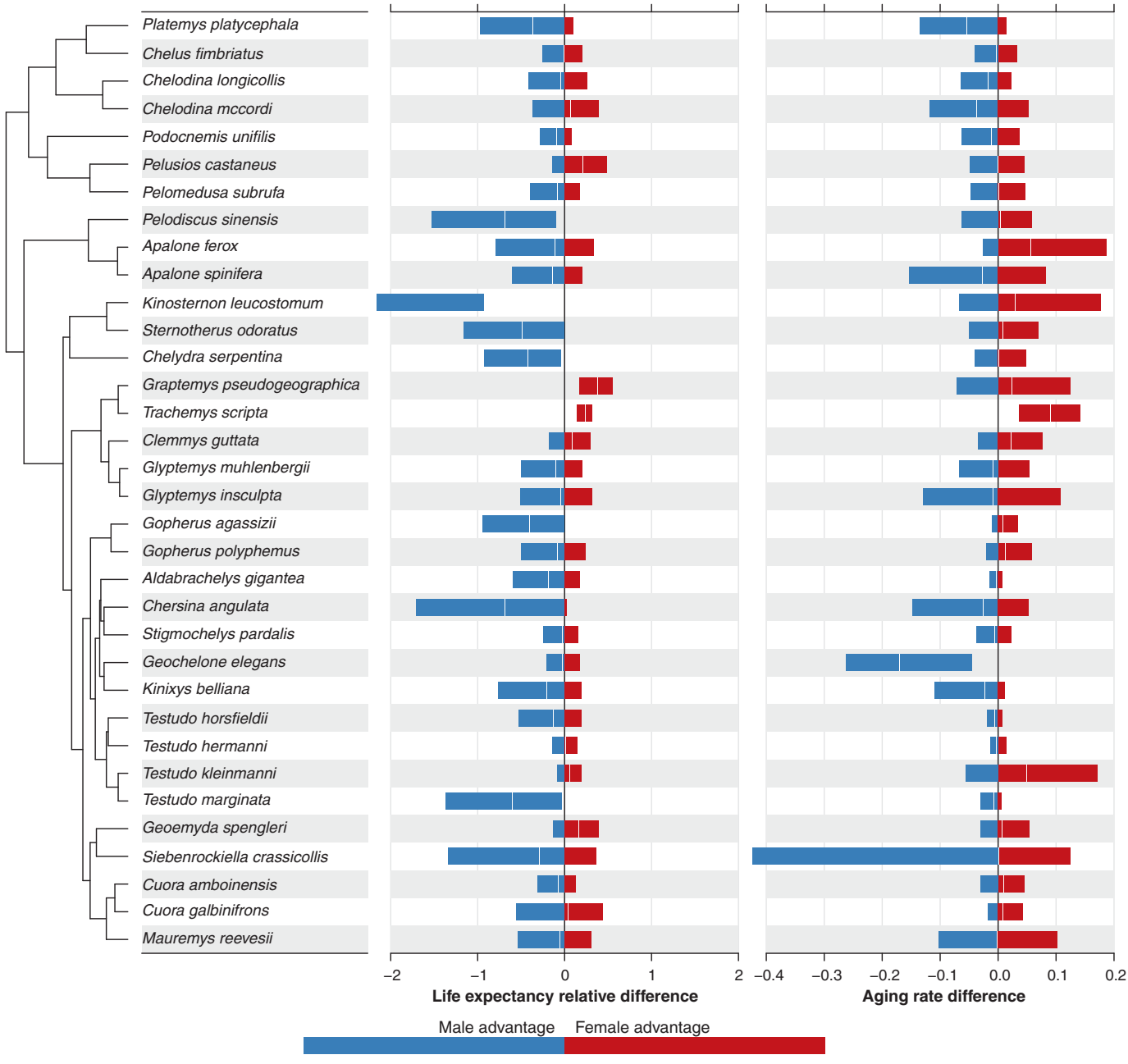


Fig. 2. Sex differences in life expectancy and aging rates. Sex differences in life expectancy were calculated as female minus male life expectancy divided by female life expectancy, whereas sex differences in aging rates were calculated as female minus male aging rates. Bars show the limits of lower and upper 95% CIs of the differences. White lines show the mean difference. Species are grouped according to their phylogenetic classification in (*35*).

results may not be reflective of aging rates and longevities in their natural environments. For instance, studies on vipers and frogs have found that populations can greatly vary their aging rates when exposed to different environmental conditions (*14*, *15*). We compared our results with data from natural environments for three species [*Chrysemys picta* (*13*), *Trachemys scripta* (*31*), and *Kinixys homeana* (*32*)]. We found considerably higher aging rates in natural environments for *T. scripta* (a = 0.04 in nature versus mean a = 0.01 in ZIMS) and *K. homeana* (a = 0.11 in nature versus mean a = 0.02 in ZIMS) (Fig. 3). The aging rates in nature of both species fell on the upper end of the posterior densities of the ZIMS aging rates and outside the 95% CIs (i.e., upper quantile for *T. scripta* of <0.0001 and for *K. homeana* of 0.007) (Fig. 3, B and F). The wild aging rate of *C. picta* was slightly lower (0.05 versus mean a = 0.08 for ZIMS), even though the wild value was contained within the 95% CIs (lower quantile of 0.15), whereas the wild population had a considerably higher baseline mortality level (Fig. 3, C and D). Still, the mechanism(s) underpinning shifts in mortality

Table 1. Bayesian PGLS for female, male, and sex differences in life expectancies and aging rates. Columns show the mean, standard error (SE), and lower and upper 95% CIs from the posterior densities of regression parameters. Zero coverage indicates to which upper or lower quantile of the posterior density zero corresponds to for a parameter: The lower the value, the more different from zero the parameter is. Dashes indicate not applicable. SSD, sexual size dimorphism; Repr. output diff., reproductive output difference.

Sex	Response	Parameters	Mean	SE	Lower	Upper	Zero coverage
Female	Life expectancy	(Intercept)	2.590	0.080	2.428	2.745	–
Female	Life expectancy	log(body weight)	0.139	0.041	0.058	0.221	0.001
Female	Life expectancy	Pagel's λ	0.083	0.063	0.004	0.237	–
Female	Aging rate	(Intercept)	0.020	0.009	0.003	0.037	–
Female	Aging rate	SSD	0.024	0.013	−0.001	0.048	0.056
Female	Aging rate	Pagel's λ	0.088	0.067	0.004	0.249	–
Male	Life expectancy	(Intercept)	2.593	0.079	2.433	2.747	–
Male	Life expectancy	log(body weight)	0.142	0.041	0.062	0.222	0.001
Male	Life expectancy	SSD	0.286	0.121	0.048	0.527	0.018
Male	Life expectancy	Pagel's λ	0.081	0.062	0.003	0.235	–
Male	Aging rate	(Intercept)	0.020	0.009	0.003	0.037	–
Male	Aging rate	SSD	0.024	0.012	−0.001	0.049	0.051
Male	Aging rate	Pagel's λ	0.087	0.064	0.004	0.240	–
Sex differences	Life expectancy	(Intercept)	−0.143	0.058	−0.258	−0.028	–
Sex differences	Life expectancy	SSD	0.231	0.084	0.063	0.400	0.006
Sex differences	Life expectancy	Pagel's λ	0.124	0.092	0.006	0.348	–
Sex differences	Aging rate	(Intercept)	−0.008	0.011	−0.030	0.015	–
Sex differences	Aging rate	Repr. output diff.	0.044	0.026	−0.007	0.095	0.094
Sex differences	Aging rate	Pagel's λ	0.105	0.085	0.004	0.319	–

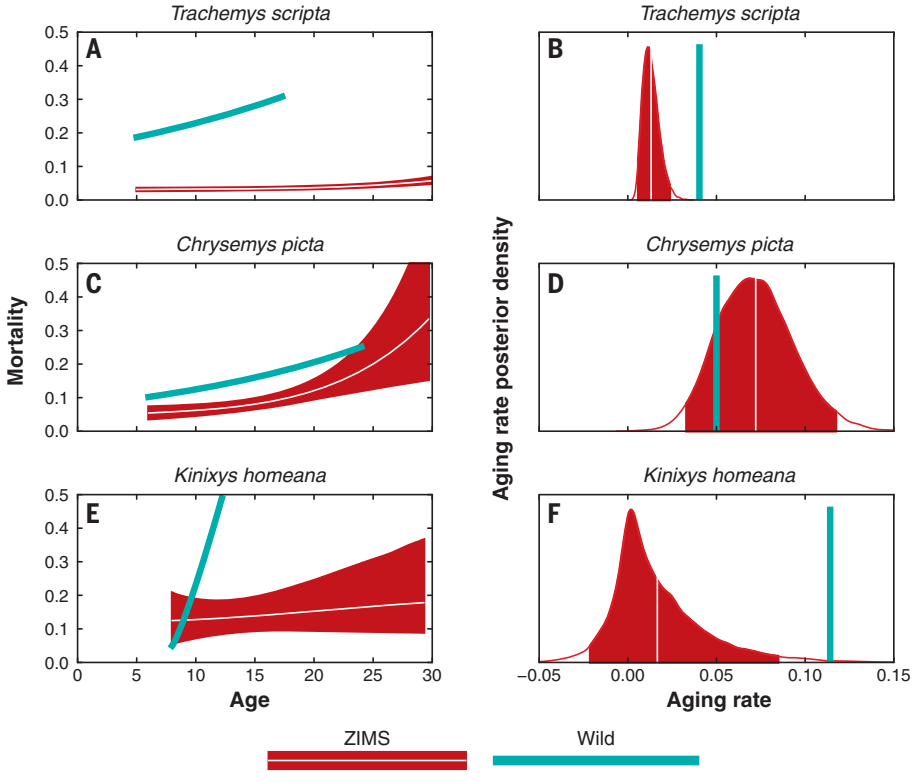


Fig. 3. Comparison of age-specific mortality trajectories and aging rates between wild and zoo and aquarium (ZIMS) female populations of three testudines species. (**A**, **C**, and **E**) Comparison of age-specific mortality trajectories. (**B**, **D**, and **F**) Comparison of aging rates. Green lines show the mortality or aging rate for the wild populations, and the red polygons show the 95% CIs of mortality or of the posterior density of aging rate for the ZIMS populations, as obtained from the BaSTA analysis. White lines depict the posterior means. Wild data for painted turtles (*C. picta*) are from (*13*), data for yellowbelly sliders (*T. scripta*) are from (*31*), and data for home's hingeback tortoise (*K. homeana*) are from (*32*).

trajectories and aging rates between wild populations and those under human care remain unclear. The disposable soma theory (*5*) posits that populations exposed to milder conditions can allocate more energy to survival rather than protection or foraging, thereby prolonging their life spans. However, this theory still predicts that senescence should be inevitable in this group.

Recent studies on humans and nonhuman primates have revealed that, in response to improvements in survival conditions, extensions in longevity emerged from a reduction in infant and juvenile mortality and in the overall level of mortality, but aging rates remained stable (*2*, *3*). Here, we show that under controlled conditions, turtles and tortoises can reduce the effects of senescence and, in some cases, even avoid them.

Current evolutionary theories of senescence have mostly been tested on birds and mammals (*33*, *34*). To fully understand how senescence molds vital rates and how environmental conditions affect senescence, further studies comparing populations under human care and natural environments are needed, particularly for underresearched tetrapods.

REFERENCES AND NOTES

1. J. Oeppen, J. W. Vaupel, *Science* **296**, 1029–1031 (2002).
2. J. M. Aburto, F. Villavicencio, U. Basellini, S. Kjærgaard, J. W. Vaupel, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5250–5259 (2020).
3. F. Colchero *et al*., *Nat. Commun.* **12**, 3666 2021).
4. T. B. L. Kirkwood, R. Holliday, *Proc. R. Soc. London B* **205**, 531–546 (1979).
5. T. B. L. Kirkwood, *Nature* **270**, 301–304 (1977).
6. C. E. Finch, *J. Gerontol. A Biol. Sci. Med. Sci.* **53A**, B235–B239 (1998).
7. O. R. Jones *et al*., *Nature* **505**, 169–173 (2014).
8. J. W. Vaupel, A. Baudisch, M. Dölling, D. A. Roach, J. Gampe, *Theor. Popul. Biol.* **65**, 339–351 (2004).
9. A. Baudisch, *Inevitable Aging? Contributions to Evolutionary-Demographic Theory* (Springer, 2008).
10. J. Gibbons, R. Semlitsch, *J. Anim. Ecol.* **51**, 523–527 (1982).
11. J. D. Congdon, R. D. Nagle, O. M. Kinney, R. C. van Loben Sels, *Exp. Gerontol.* **36**, 813–827 (2001).
12. J. D. Congdon *et al*., *Exp. Gerontol.* **38**, 765–772 (2003).
13. D. A. Warner, D. A. W. Miller, A. M. Bronikowski, F. J. Janzen, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6502–6507 (2016).
14. T. Tully, J.-F. Le Galliard, J.-P. Baron, *J. Anim. Ecol.* **89**, 2704–2716 (2020).
15. H. Cayuela *et al*., *J. Anim. Ecol.* 10.1111/1365-2656.13545 (2021).
16. Species360 Zoological Information Management System (ZIMS) (2020); www.Species360.org.
17. F. Colchero, J. S. Clark, *J. Anim. Ecol.* **81**, 139–149 (2012).
18. F. Colchero, O. R. Jones, M. Rebke, *Methods Ecol. Evol.* **3**, 466–470 (2012).
19. Materials and methods are available as supplementary materials online.
20. J.-M. Gaillard *et al*., in *Encyclopedia of Evolutionary Biology*, R. M. Kliman, Ed. (Elsevier, 2016), pp. 312–323.
21. J. F. Berry, R. Shine, *Oecologia* **44**, 185–191 (1980).
22. L. A. Hoekstra, T. S. Schwartz, A. M. Sparkman, D. A. W. Miller, A. M. Bronikowski, *Funct. Ecol.* **34**, 38–54 (2020).
23. F. Eckhardt, P. M. Kappeler, C. Kraus, *Sci. Rep.* **7**, 11397 (2017).
24. J.-F. Lemaître *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8546–8553 (2020).
25. R. Shine, *Am. Nat.* **138**, 103–122 (1991).
26. T. Madsen, R. Shine, *Oecologia* **92**, 40–47 (1992).
27. G. Stark, K. Tamar, Y. Itescu, A. Feldman, S. Meiri, *Biol. J. Linn. Soc.* **125**, 730–740 (2018).
28. T. H. Clutton-Brock, K. Isvaran, *Proc. Biol. Sci.* **274**, 3097–3104 (2007).
29. D. Venegas *et al*., *Gen. Comp. Endocrinol.* **236**, 35–41 (2016).
30. L. Zhang, X. Lu, *Evol. Biol.* **40**, 84–91 (2013).
31. J. W. Gibbons, *Bioscience* **37**, 262–269 (1987).
32. H. Cayuela *et al*., *Biol. J. Linn. Soc.* **128**, 251–262 (2019).
33. O. R. Jones *et al*., *Ecol. Lett.* **11**, 664–673 (2008).
34. D. H. Nussey, H. Froy, J.-F. Lemaitre, J.-M. Gaillard, S. N. Austad, *Ageing Res. Rev.* **12**, 214–225 (2013).
35. A. G. Pereira, J. Sterli, F. R. R. Moreira, C. G. Schrago, *Mol. Phylogenet. Evol.* **113**, 59–66 (2017).
36. testudinesSenescenceRcode, Github (2022); https://github.com/fercol/testudinesSenescenceRcode.

ACKNOWLEDGMENTS

We dedicate this manuscript to the late James W. Vaupel in recognition of his invaluable support. We thank the recordkeepers and the more than 1000 institutions that are part of Species360 for providing data. We thank L. Bland for her editorial help. **Funding:** Institutional support was provided by the University of Southern Denmark. This work was supported by a research grant (34439) from VILLUM FONDEN, awarded to R.d.S., and the National Institute on Aging grant P01 AG031719 (primary investigator: J. W. Vaupel; supporting: F.C. and D.A.C.). **Author contributions:** Conceptualization: F.C., R.d.S., D.A.C., and A.B. Methodology: F.C. Investigation: R.d.S. and F.C. Visualization: R.d.S. and F.C. Funding acquisition: R.d.S. Project administration: F.C. Supervision: F.C. Writing – original draft: R.d.S., F.C., and A.B. Writing – review & editing: R.d.S., F.C., D.A.C., and A.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Raw data used for survival analysis to estimate life expectancies and aging rates (Species360 Data Use Approval no. 91356) cannot be publicly shared, as Species360 is the custodian (not the owner) of their members' data. Raw data are accessible through Research Request applications (form available at https://docs.google.com/forms/d/1znoy62VEkDlhAp_0RfEvF7Zsx03g4W5AlppJHqo3_WQ/viewform?edit_requested=true&pli=1). All data needed to reproduce the conclusions of the Bayesian PGLS regressions are present in the main text or the supplementary materials. All R code necessary to reproduce the results is available at (*36*).
License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl7811
Materials and Methods
Figs. S1 and S2
Tables S1 to S7
References (*37*–*167*)
MDAR Reproducibility Checklist

Submitted 6 August 2021; accepted 28 April 2022
10.1126/science.abl7811

DNA DAMAGE REPAIR

Small-molecule activation of OGG1 increases oxidative DNA damage repair by gaining a new function

Maurice Michel[1]*†, Carlos Benítez-Buelga[1,2]†, Patricia A. Calvo[3]‡, Bishoy M. F. Hanna[1]‡, Oliver Mortusewicz[1]‡, Geoffrey Masuyer[4,5]‡, Jonathan Davies[5]‡, Olov Wallner[1]‡, Kumar Sanjiv[1]‡, Julian J. Albers[1], Sergio Castañeda-Zegarra[1,6], Ann-Sofie Jemth[1], Torkild Visnes[7], Ana Sastre-Perona[8], Akhilesh N. Danda[1], Evert J. Homan[1], Karthick Marimuthu[1], Zhao Zhenjun[1], Celestine N. Chi[9], Antonio Sarno[10], Elisée Wiita[1], Catharina von Nicolai[1], Anna J. Komor[11], Varshni Rajagopal[1], Sarah Müller[1], Emily C. Hank[1], Marek Varga[1], Emma R. Scaletti[5,12], Monica Pandey[1,13], Stella Karsten[1], Hanne Haslene-Hox[7], Simon Loevenich[7], Petra Marttila[1], Azita Rasti[1], Kirill Mamonov[1], Florian Ortis[1], Fritz Schömberg[14], Olga Loseva[1], Josephine Stewart[1], Nicholas D'Arcy-Evans[1], Tobias Koolmeister[1], Martin Henriksson[1], Dana Michel[15], Ana de Ory[16], Lucia Acero[8], Oriol Calvete[17], Martin Scobie[1], Christian Hertweck[11,18], Ivan Vilotijevic[14], Christina Kalderén[1], Ana Osorio[17,19], Rosario Perona[2,19], Alexandra Stolz[20], Pål Stenmark[5,12], Ulrika Warpman Berglund[1]§, Miguel de Vega[3], Thomas Helleday[1,13]*

Oxidative DNA damage is recognized by 8-oxoguanine (8-oxoG) DNA glycosylase 1 (OGG1), which excises 8-oxoG, leaving a substrate for apurinic endonuclease 1 (APE1) and initiating repair. Here, we describe a small molecule (TH10785) that interacts with the phenylalanine-319 and glycine-42 amino acids of OGG1, increases the enzyme activity 10-fold, and generates a previously undescribed β,δ-lyase enzymatic function. TH10785 controls the catalytic activity mediated by a nitrogen base within its molecular structure. In cells, TH10785 increases OGG1 recruitment to and repair of oxidative DNA damage. This alters the repair process, which no longer requires APE1 but instead is dependent on polynucleotide kinase phosphatase (PNKP1) activity. The increased repair of oxidative DNA lesions with a small molecule may have therapeutic applications in various diseases and aging.

The well-established role of 8-oxoguanine (8-oxoG) DNA glycosylase 1 (OGG1) in DNA repair, cancer, and inflammation led to the development of OGG1 inhibitors—among them O8, SU0268, and TH5487—with potential uses as anticancer or anti-inflammatory agents (*1*–*3*). Around the same time, small-molecule activators of OGG1 were reported (*4*, *5*). Applications of these molecules for targeting OGG1 in high–oxidative stress conditions, such as Alzheimer's disease and obesity, have been suggested (*6*–*8*). In the literature, prominent examples of enzyme activators exhibit their effect by allosteric control (*9*–*12*). In the case of OGG1, the excised 8-oxoG nucleobase has been reported to assist catalysis (*13*), and we previously identified potential secondary druggable sites in OGG1 (*14*). Importantly, whereas OGG1 has only a weak β-elimination function, other DNA glycosylases are bifunctional and process apurinic or apyrimidinic sites (AP sites) using a β,δ-lyase activity (Fig. 1A) (*15*). Thus, to be able to apply small-molecule activators of OGG1 in disease models, we sought to investigate the mode of action at a molecular level.

We synthesized previously reported compounds (*4*) (see supplementary materials for details) and assessed their ability to thermally stabilize OGG1 using differential scanning fluorimetry (DSF) (fig. S1). We found that TH10785, which is compound C in (*4*), stabilized OGG1 by 3.8° ± 0.7°C (100 μM). Next, we measured in vitro OGG1 activity using a fluorescence-based biochemical assay (*3*) in the presence of TH10785 or apurinic endonuclease 1 (APE1) to increase the reaction rate in a coupled enzyme assay (Fig. 1B). Surprisingly, TH10785 alone, in the absence of APE1, increased the reaction rate in a concentration-dependent fashion and reached a maximal enhancement at 6.25 μM. At higher concentrations, the turnover progressively decreased to the levels of the reaction rate of the dimethyl sulfoxide (DMSO) control, producing a bell-shaped activity profile (fig. S2). This increase in activity was specific for OGG1 and its substrate (fig. S3 and table S1). To test if TH10785 stimulates the OGG1 β-lyase activity on AP sites, we performed the assay using a uracil-containing substrate in the presence of human uracil-DNA glycosylase 2 (UNG2). The AP sites generated by UNG2 were resolved with increasing efficiency as well (fig. S4), suggesting a stimulation of the residual β-lyase activity of OGG1. To confirm this, we analyzed the reaction products that were generated after in vitro incubation of OGG1 with a radioactively labeled double-stranded DNA substrate containing an internal 8-oxoG:C pair or an AP site. In the presence of TH10785, the β-lyase activity of OGG1 was stimulated for both substrates (Fig. 1, C and D). To exclude an effect on the DNA glycosylase activity of OGG1, we measured the number of AP sites generated in the presence of TH10785 at 6.25 μM. No effect was observed (fig. S5), suggesting that OGG1 activation by TH10785 is confined to its lyase activity. Unexpectedly, the presence of TH10785 yielded two products, as observed using gel electrophoresis. One corresponds to the expected 18mer-3′-phosphate unsaturated aldehyde (PUA) and the second to a higher electrophoretic mobility (Fig. 1, C and D). Investigating the product composition (Fig. 1E) (*16*), we observed that neither of the products were elongated by adding DNA polymerase X (*17*) with deoxyguanosine triphosphate (dGTP) (lane d, PolX + dGTP), indicating the absence of a 3′-OH end. However, after treating the degradation products with T4 polynucleotide kinase, which has a robust 3′-phosphatase activity, the electrophoretic fastest-migrating product was converted to the more slowly migrating 18mer-3′OH product (lane g). This product was able to be elongated by DNA polymerase X with dGTP, yielding a 19mer species (lane h). Complete repair to the original 34mer was obtained by adding T4 DNA ligase (lane i). This in vitro pathway reconstitution demonstrates that TH10785 allows OGG1 to efficiently cleave AP sites through a previously undescribed β-δ activity.

To understand how an OGG1 β,δ-lyase activity is controlled by TH10785, we solved the cocrystal structure of TH10785 bound to OGG1, which confirmed binding to the catalytic site (Fig. 1F and fig. S6). Additionally, we confirmed the need of Lys^{249} to form a OGG1-abasic site intermediate adduct (fig. S7). Complementary

[1]Science for Life Laboratory, Department of Oncology-Pathology, Karolinska Institutet, 171 76 Stockholm, Sweden. [2]Instituto de Investigaciones Biomédicas Alberto Sols (CSIC/UAM), 28029 Madrid, Spain. [3]Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), 28049 Madrid, Spain. [4]Department of Pharmacy and Pharmacology, Centre for Therapeutic Innovation, University of Bath, Bath BA2 7AY, UK. [5]Department of Biochemistry and Biophysics, Stockholm University, 106 91 Stockholm, Sweden. [6]Department of Clinical and Molecular Medicine (IKOM), Norwegian University of Science and Technology, 7491 Trondheim, Norway. [7]Department of Biotechnology and Nanomedicine, SINTEF Industry, N-7465 Trondheim, Norway. [8]Experimental Therapies and Novel Biomarkers in Cancer, Hospital La Paz Institute for Health Research (IdiPAZ), Madrid, Spain. [9]Disciplinary Domain of Medicine and Pharmacy, Faculty of Medicine, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden. [10]Department of Environment and New Resources, SINTEF Ocean, N-7496 Trondheim, Norway. [11]Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Department of Biomolecular Chemistry, 07745 Jena, Germany. [12]Department of Experimental Medical Science, Lund University, Lund, Sweden. [13]Sheffield Cancer Centre, Department of Oncology and Metabolism, University of Sheffield, Sheffield S10 2RX, UK. [14]Institute of Organic and Macromolecular Chemistry, Friedrich Schiller University Jena, 07743 Jena, Germany. [15]Chemical Processes and Pharmaceutical Development, Unit Process Chemistry I, Research Institutes of Sweden – RISE, 151 36 Södertälje, Sweden. [16]Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University, 106 91 Stockholm, Sweden. [17]Familial Cancer Clinical Unit, Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain. [18]Institute of Microbiology, Friedrich-Schiller-University Jena, 07743 Jena, Germany. [19]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, 28029 Madrid, Spain. [20]Institute of Biochemistry II and Buchmann Institute for Molecular Life Science, Goethe University Frankfurt, 60590 Frankfurt, Germany.
***Corresponding author. Email: maurice.grube@scilifelab.se (M.M.); thomas.helleday@scilifelab.se (T.H.)**
†These authors contributed equally to this work. ‡These authors contributed equally to this work. §Present address: Oxcia AB, 113 34 Stockholm, Sweden.

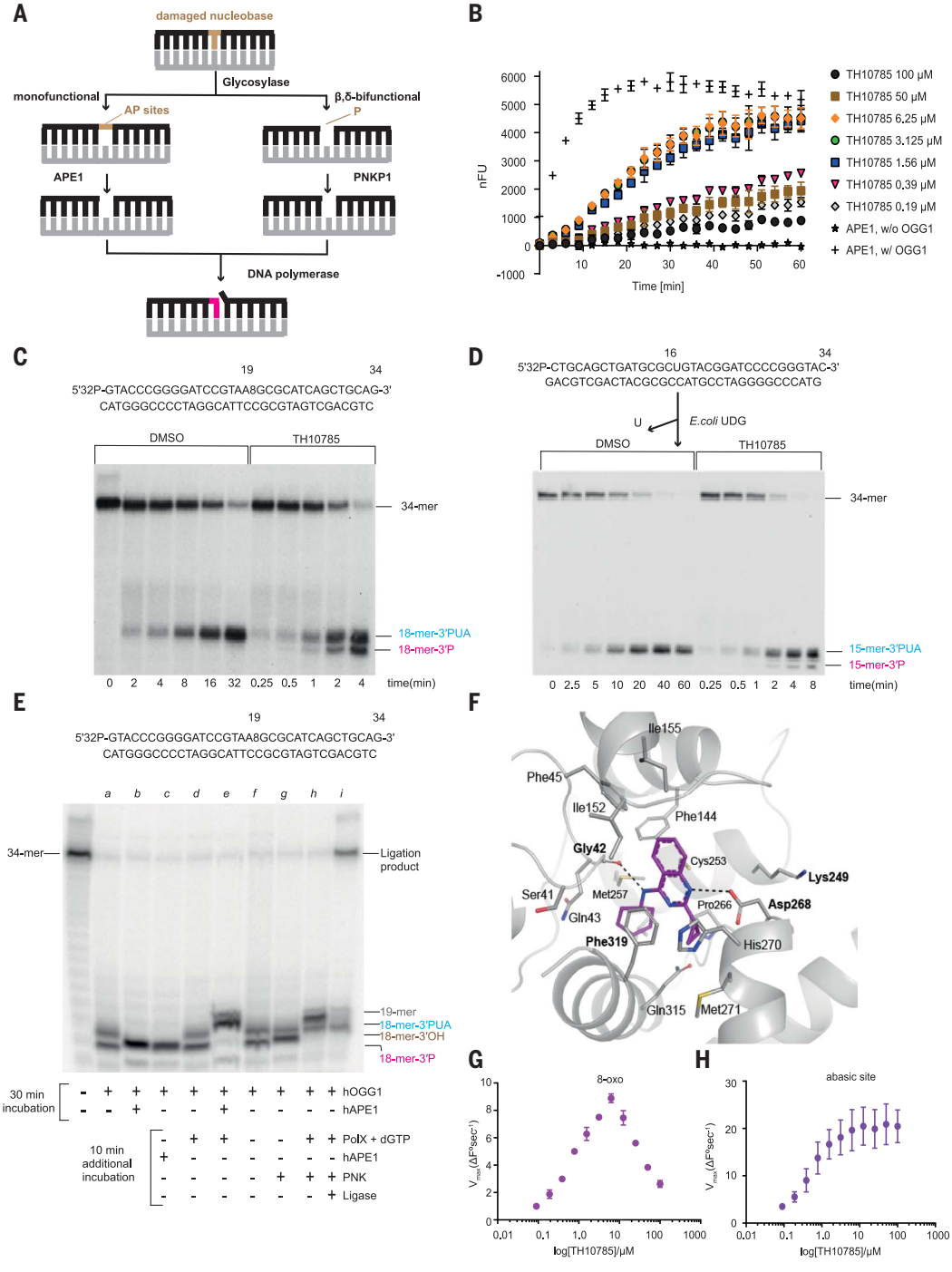


Fig. 1. OGG1 activator TH10785 induces a de novo β,δ-elimination in vitro, allowing for AP sites as new substrates. (**A**) The repair fate for substrates of monofunctional and β,δ-bifunctional DNA glycosylases is either APE1 or PNKP1 dependent. OGG1 as a bifunctional glycosylase with residual product-assisted β-elimination activity recognizes and excises 8-oxoG. Because of the inefficient process, APE1 performs incision of the AP site and excision of PUA to generate a 3′-hydroxyl end compatible with a DNA polymerase. Other DNA glycosylases, such as endonuclease VIII–like proteins 1 and 2 (NEIL1/2), can catalyze an additional β,δ-elimination in addition to the DNA glycosylase function, which includes incision of the AP site and removal of resulting PUA. This generates a 3′-hydroxyl end, which is only compatible with a DNA polymerase after additional elimination of the generated 3′-phosphate by PNKP1. Thus, these enzymes are functioning independently of APE1. (**B**) TH10785 dose-dependently stimulates the excision of 8-oxoA:C by OGG1 in the absence of APE1. Data are presented as averages ± SD of two technical replicates from two independent experiments. nFU, normalized fluorescence units generated after subtracting the DMSO control. (**C**) A radioactively labeled 8oxoG:C–containing DNA substrate confirms an AP-lyase activity of OGG1 but yields two distinct products. In the assay, 1 nM of the [^{32}P]5′-labeled 8oxoG–containing substrate was incubated with 25 nM OGG1 and either 10% DMSO or 6.25 μM TH10785, as indicated. Samples were analyzed by 7M urea–20% PAGE (polyacrylamide gel electrophoresis) and autoradiography. The positions of the products are as indicated. (**D**) Using 1 nM U:C–containing DNA substrate and 6 nM *Escherichia coli* uracil-DNA glycosylase (UDG) in the same assay setup confirms that AP sites are substrates and yields two distinct products. (**E**) APE1-independent de novo β,δ-elimination of AP sites by OGG1 is confirmed in the presence of TH10785. In the assay, 1 nM of the 8oxoG-containing substrate was incubated with 6.25 μM TH10785 and 50 nM OGG1 either in the presence (+) or absence (−) of 0.2 nM APE1. After incubation for 30 min at 37°C, reactions were further incubated in the presence of the indicated protein. The position and a scheme of the reaction products are shown. PNK, polynucleotide kinase. (**F**) Close-up view of ligand TH10785 (magenta) binding to human OGG1. Important residues in the binding site are marked, and hydrogen-bond interactions are shown as black dashed lines. A central π-stacking of the quinazoline ring with Phe^{319} was observed, supporting the inactivity of the Phe^{319}→Ala (Phe319Ala) mutant with TH10785. Further stabilization of the complex is provided by a hydrogen bond between the proton of the secondary amine and Gly^{42}. In addition, the cyclopropyl group is located within the shallower lipophilic pocket surrounded by Pro^{266} and Met^{271}. Notably, in the hOGG1 complex, TH10785 makes a hydrogen bond with the acidic side chain of Asp^{268}, whereas Lys^{249} is oriented away from the catalytic pocket. (**G** and **H**) Values for the kinetic parameter v_{max} determined from saturation curves with (G) 8-oxoA–containing and (H) AP site–containing substrates reflect an increased turnover of AP site substrate in the presence of TH10785. Although binding to the active site occurs, no competitive behavior is observed for AP sites. By contrast, at saturating concentrations, TH10785 competes with the 8-oxoA–containing substrate. Data are represented as mean values ± SD. Data are the average of two independent experiments.

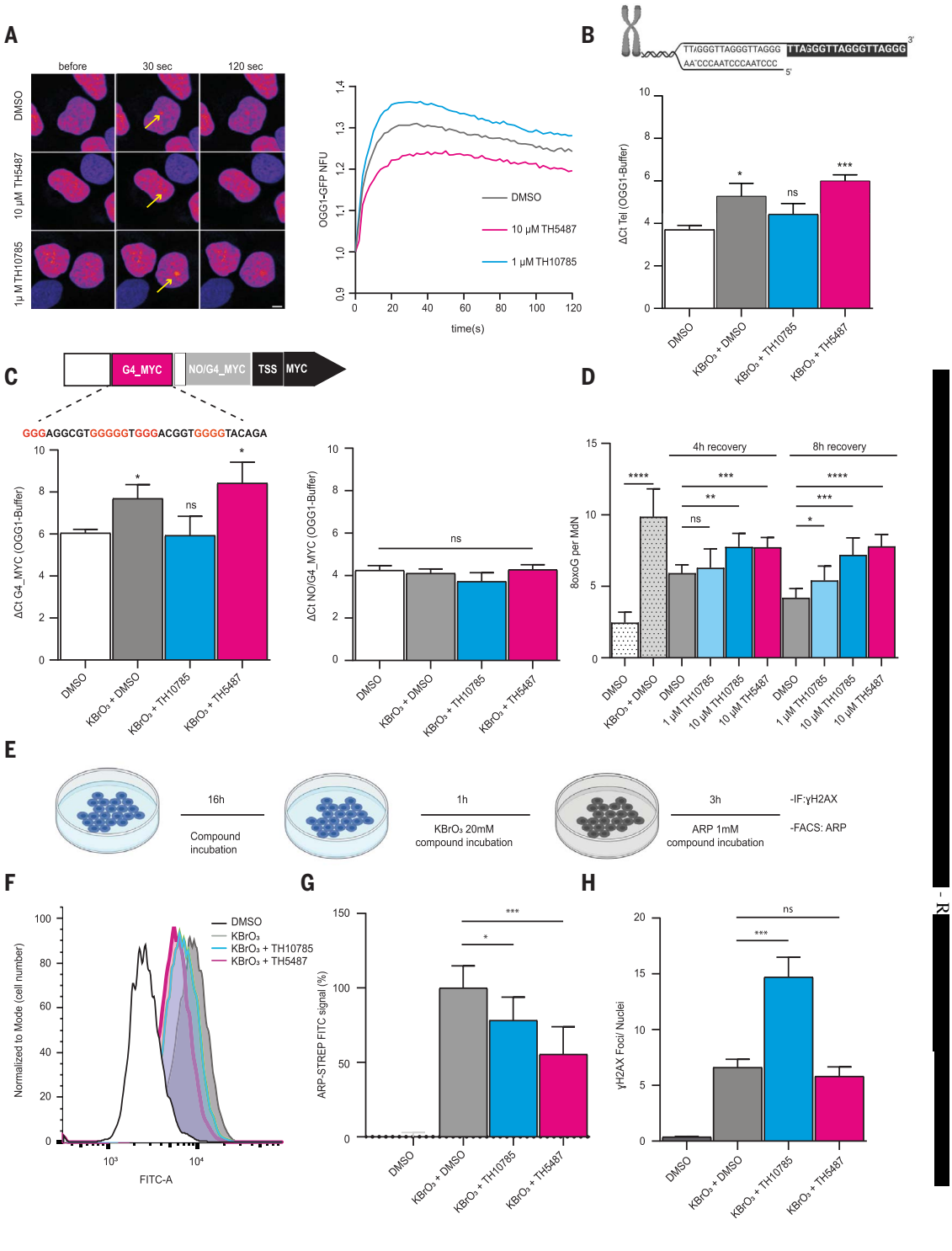


Fig. 2. TH10785 allows OGG1 to increase DNA repair by addressing AP sites. (**A**) In the presence of TH10785, more OGG1 is recruited to laser-damaged sites. Shown on the left are representative live-cell confocal images of microirradiated U2OS OGG1-GFP cells treated with OGG1 inhibitor (TH5487, 10 µM), OGG1 activator (TH10785, 1 µM), or vehicle (DMSO). The yellow arrows show the microirradiated position. Shown on the right are connecting lines (±SEM) of OGG1-GFP recruitment kinetics. (The complete set of images as well as a comparative analysis of maximum OGG1-GFP intensity at laser-irradiated sites are represented in fig. S22.) Scale bar, 5 µm. NFU, normalized fluorescence units. (**B** and **C**) TH10785-mediated repair by OGG1 decreases oxidative damage in guanine-rich regions of the genome. Results for the OGG1 incision assay on DNA extracted from U2OS cells treated with 20 mM $KBrO_3$ for 1 hour alone or in combination with TH10785 (10 µM), TH5487(10 µM), or DMSO are shown. Each bar represents the mean ± SEM of (B) telomere delta cycle threshold (ΔCt Tel), (C, left) G4_MYC ΔCt, and (C, right) control NO/G4_MYC ΔCt. ΔCt values are calculated as target Ct for the DNA incubated with OGG1-target Ct for the DNA incubated with the buffer. Data are the average of three independent experiments with three or four replicates. Statistical significance was calculated using the Mann-Whitney test. ns, nonsignificant; $^{*}P < 0.05$; $^{***}P < 0.001$. (**D**) With TH10785 present, accumulated 8-oxodG in U2OS cells after oxidative stress declines slower than in the DMSO but faster than in the TH5487 control. Cells were treated with 20 mM $KBrO_3$ for 1 hour, and afterward, cells were exposed to 1 µM or 10 µM TH10785, 10 µM TH5487, or DMSO at the indicated doses for 4 and 8 hours. The amount of genomic 8-oxodG was quantified with LC-MS/MS. Each bar represents the mean ± SD of three replicates from two independent experiments. Statistical significance was calculated using an unpaired two-tailed Student's *t* test. ns, nonsignificant; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$. MdN, million deoxynucleotides. (**E** to **H**) TH10785 increases AP site cleavage during base excision repair initiation. The scheme of treatment and incubation with the ARP_probe in U2OS cells is shown in (E). A histogram overlay for the ARP-STREP_FITC geometric mean of cells at different treatment conditions is shown in (F). A comparative analysis of ARP-STREP_FITC signal induction over DMSO, reported in percentage, is shown in (G). Each bar represents the mean ± SD. Data are the average of two independent experiments with three or four biological replicates. Statistical significance was calculated using an unpaired two-tailed Student's *t* test. ns, nonsignificant; $^{*}P < 0.05$; $^{***}P < 0.001$. In (H), a comparative analysis of the average number of γH2AX foci per nucleus using quantitative microscopy is shown. Each bar represents the mean foci per nuclei ± SEM of five independent experiments. For each experiment, a minimum of 200 cells were analyzed. Statistical analysis was performed using the Mann-Whitney test. ns, nonsignificant; $^{***}P < 0.001$. Mean ± SEM values for Fig. 2A and Fig. 2, B, C, and H, are reported in tables S8 and S9, respectively.

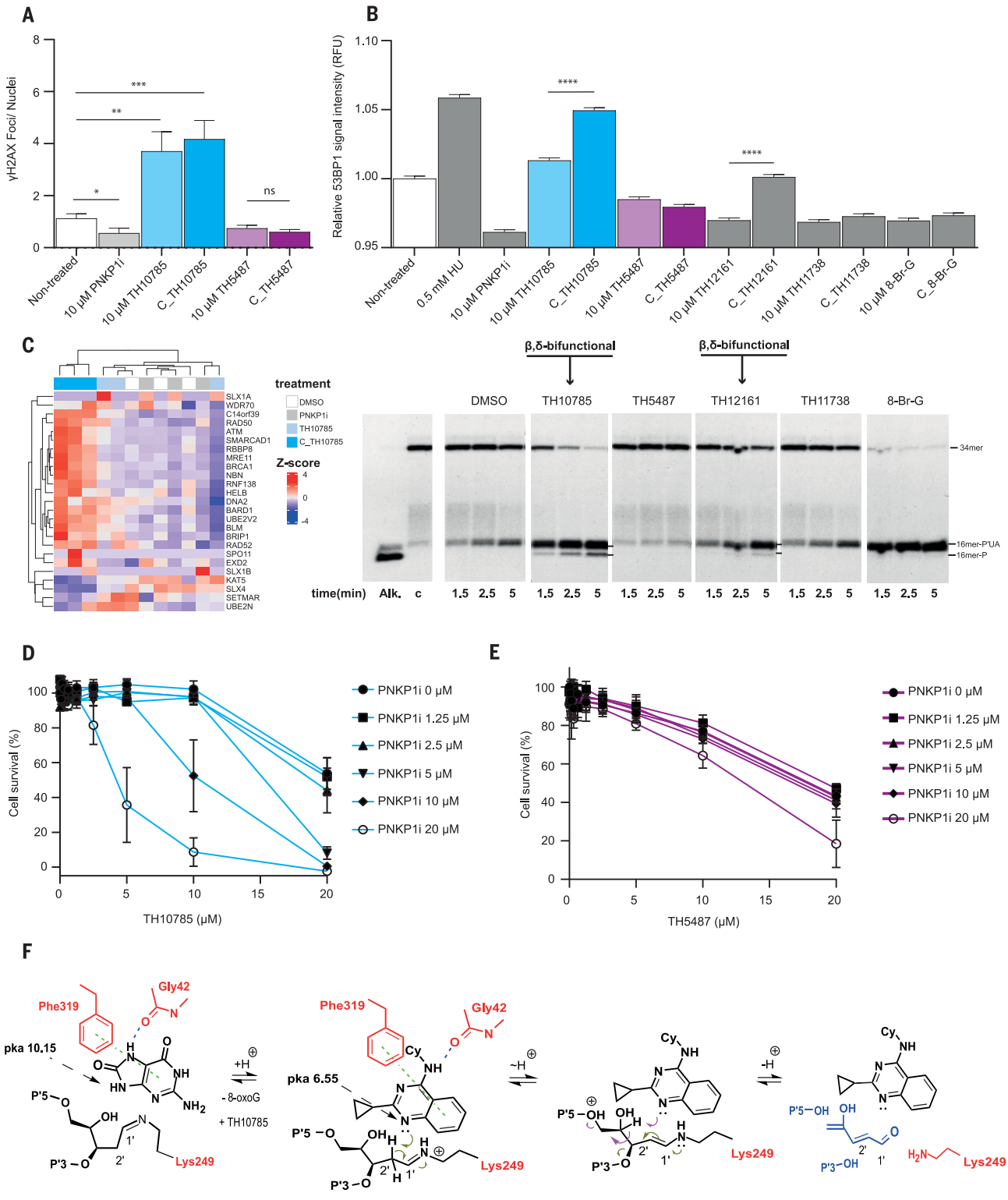


Fig. 3. TH10785-induced OGG1 β,δ-lyase activity shifts cells toward PNKP1 dependence. (**A** and **B**) OGG1 repair products cause DNA damage when β,δ-elimination is established. Quantification of γH2AX foci per nuclei is shown in (A). Each bar represents the mean ± SEM of three independent experiments. Statistical analysis was performed using the Mann-Whitney test. ns, nonsignificant; *P < 0.05; **P < 0.001; ***P < 0.0001. A quantification of 53BP1 nuclear intensity in U2OS cells treated for 24 hours with the indicated conditions is shown at the top of (B). The notation "C_" means a combination of the indicated TH molecule and PNKP1i. Each bar represents the mean ± SEM from three independent experiments in U2OS cells treated for 24 hours with the indicated conditions. For each experiment, a minimum of 200 cells were analyzed. Statistical analysis was performed using the Mann-Whitney test. ****P < 0.0001. The bottom of (B) shows a duplicated image from fig. S20 to illustrate that OGG1 activators that cause β,δ-lyase activity in vitro induce DDR activation when combined with PNKP1i in cells. Alk., alkaline DNA cleavage control; c, control without enzyme. (**C**) OGG1 activation through

β,δ-elimination in combination with PNKP1 inhibition causes an up-regulation of members of the DDR. A heat map showing the expression of genes involved in "DNA double-strand break procession" (Gene Ontology gene set) is shown. U2OS cells were treated with 10 μM TH10785, 10μM PNKP1i, 10μM TH10785 with 10 μM PNKP1i, or DMSO for 24 hours (biological replicates n = 3). Color indicates row-wise z-scored DESeq2-normalized counts. Samples and genes are clustered hierarchically (Euclidean distance, complete linkage). (**D** and **E**) OGG1 activation through β,δ-elimination in combination with PNKP1 inhibition decreases cell survival. Connecting lines show mean ± SEM normalized viability values of U2OS cells treated for 72 hours with the indicated doses of PNKP1i in combination with (D) TH10785 or (D) TH5487. Data are the average of three independent experiments. (**F**) Molecular mechanism of OGG1 enzyme-directed small-molecule catalysis: TH10785 establishes active-site binding of OGG1 by addressing key amino acid residues Phe^{319} (green dashed line, π-stacking) and Gly^{42} (blue dashed line, H-bonding). An appropriately placed nitrogen base enables abstraction of the 2′C-proton and increases β-elimination. Further, a de novo δ-elimination is installed by the catalyst, possibly occurring simultaneously to product release. Mean ± SEM values for Fig. 3A and Fig. 3, B, D, and E, are reported in tables S9 and S10, respectively.

to this, we confirmed active-site binding by producing relevant enzyme mutants and assessed their activity in the presence of TH10785 (fig. S8). Together, these results demonstrate that catalytic-site binding is critical for OGG1 activation by TH10785. To determine how OGG1 activators can engage the active site and not inhibit the occurring enzymatic reactions, we performed molecular dynamics simulations (fig. S9 and movies S1 and S2) and dissociation constant (K_D) measurements using fluorescence spectroscopy (table S2) (*18*). Both indicated that the affinity of TH10785 to OGG1 is rather low (table S2). The affinity of TH10785 to OGG1 (K_D = 5.5 ± 1.7 μM) increased when adding an AP site analog containing double-stranded DNA (K_D = 1.3 ± 0.3 μM). To investigate this apparent preference for a ternary complex of DNA, OGG1, and TH10785, we determined the kinetic parameter of the corresponding reactions. Using the AP site substrate, the enzymatic turnover, v_{max}, increased dose dependently to 20-fold compared with that of the lowest concentration (Fig. 1, G and H, and fig. S10). For the 8-oxoadenine (8-oxoA) substrate, the enzymatic turnover increased as well but followed the aforementioned bell-shaped curve, going through a maximum at 6.25 μM of TH10785. The results of a DNA binding–affinity assay further showed that at the relevant concentration of 6.25 μM, TH10785 does not compete with 8-oxo-deoxyguanosine (8-oxodG) for binding to OGG1, unlike the OGG1 inhibitor TH5487 (fig. S11). Collectively, this indicates that TH10785 acts as an essential cofactor by a de facto increased enzyme activity through stimulation of substrate turnover. As evident from the kinetic analysis, higher concentrations of TH10785 compete with the nucleobase-containing substrate, but not with the AP site–containing substrate, for binding to OGG1 (Fig. 1, G and H).

These results led us to the hypothesis that TH10785 acts in a similar manner as 8-oxoG in the product-assisted catalysis by 8-oxoG that was postulated by Fromme *et al.* (*13*). To test this, we built a ternary complex of OGG1, substrate DNA, and TH10785, which was docked in the emptied binding pocket using induced-fit simulation (figs. S12 and S13 and movie S3). Using this model, we hypothesized the 1-N position to be the center of a proton abstraction at the C′2 position of the open-chain deoxyribose. To investigate this, we systematically evaluated a structure activity relationship based on activation potential [median activation concentration (AC_{50})], stabilization in DSF, pK_a, K_D, and product composition in the electromobility shift assay used above, where K_a is the acid dissociation constant (supplementary text and figs. S14 to S20). Based on the results that we obtained, we concluded that TH10785 acts as a catalyst, enhancing β-elimination and inducing a previously undescribed δ-elimination by OGG1. By binding into the OGG1 active site with the DNA substrate bound via a transient Schiff-Base intermediate, TH10785 provides an appropriately positioned nitrogen base to mediate proton abstraction. As a consequence of this well-balanced affinity and activity, induction of a β,δ-elimination cascade is established.

To determine whether TH10785 also increases OGG1 activity in cells, we first confirmed target engagement to OGG1 using the cellular thermal shift assay (*19*) (fig. S21). Next, using live-cell confocal microscopy, we measured OGG1–green fluorescent protein (GFP) recruitment kinetics to laser-microirradiated sites. Whereas TH5487 reduced recruitment, TH10785 increased OGG1-GFP recruitment to laser-induced DNA damage sites (Fig. 2A and fig. S22). We hypothesized that TH10785-mediated OGG1 recruitment is controlled by increased turnover of oxidative DNA damage repair through the previously uncharacterized AP-lyase function. Consequently, a quicker product release would occur. This is strengthened by the observed higher OGG1 mobility using fluorescence recovery after photobleaching and the absence of an effect of TH10785 toward the DNA binding capacity of OGG1 (figs. S23 and S11, respectively). To test this, we challenged U2OS cells with $KBrO_3$ for 1 hour and measured the accumulation of OGG1 substrates by quantitative PCR (qPCR) at regions prone to oxidation such as the telomeres (*20*) or a G4 quadruplex–forming sequence at the *MYC* promoter (G4_MYC) (*21*). As a negative control, a region within the *MYC* promoter that lies outside the G4 quadruplex–forming sequence was included (NO/G4_MYC). At the control region, we found unchanged levels of oxidative DNA damage. By contrast, both telomeric and G4_MYC regions accumulated oxidative DNA damage in response to $KBrO_3$, and the corresponding levels were either reduced or increased by TH10785 or TH5487 treatments, respectively (Fig. 2, B and C). We performed a modified comet assay and, in agreement with this data, obtained reduced tail moments with TH10785 present (fig. S24).

Because $KBrO_3$ generates both 8-oxoG and AP sites, qPCR or Comet assays do not allow discrimination of glycosylase and AP-lyase activities. Therefore, we quantified 8-oxodG levels in U2OS cells exposed to $KBrO_3$ for 1 hour by liquid chromatography with tandem mass spectroscopy (LC-MS/MS) and immunofluorescence (IF). After treatment, we monitored the resulting levels of 8-oxodG at different time points (4 and 8 hours in Fig. 2D or 3 and 7 hours in fig. S25). We found declining 8-oxodG levels over time when using different TH10785 concentrations. By contrast, inhibition of the glycosylase activity was observed with TH5487. Considering that TH10785 is not a potent OGG1 DNA glycosylase inhibitor (fig. S5), decreased repair of 8-oxodG in presence of TH10785, compared with DMSO, may reflect the redirection of OGG1 function toward resolving AP sites or competition for 8-oxoG substrates. To test this hypothesis, we treated U2OS cells with $KBrO_3$ for 1 hour, followed by incubation with an aldehyde reactive probe (1 mM for 3 hours) (*22*) to label AP sites and monitor by fluorescence-activated cell sorting (FACS) (Fig. 2, E and F). Although both OGG1 modulators reduced $KBrO_3$-induced AP sites (Fig. 2G), we found a reduction of nuclear γH2AX foci formation (a marker of DNA strand breaks) with TH5487 (Fig. 2H), demonstrating that impairment of OGG1 glycosylase activity leads to a reduced number of AP sites. By contrast, we an induction of nuclear γH2AX foci formation with TH10785 (Fig. 2H), confirming the generation of DNA strand breaks by the catalytic activity of TH10785 in cells. Collectively, these results point toward a new cellular role of TH10785-activated OGG1: favoring AP sites over 8-oxoG.

Next, we tested the extent to which TH10785 can induce β,δ-elimination in cells. We hypothesized that simultaneous stimulation of β,δ-elimination and blocking of polynucleotide kinase phosphatase (PNKP1) activity should overload the system with unrepaired DNA single-strand breaks (Fig. 1A). Thus, the DNA damage response (DDR) was measured by IF using the markers γH2AX as well as 53BP1 in

U2OS cells exposed to OGG1 inhibitor or activator (Fig. 3A and fig. S26) and analog compounds (Fig. 3B and table S6) alone or in combination with PNKP1 inhibitor (PNKP1i) (*23*). We found that PNKP1i only induced a strong DDR in combination with OGG1 activators, causing β,δ-lyase activity in vitro. To assess this causality, we monitored transcriptional changes using RNA sequencing and found that PNKP1i in combination with TH10785, but not the single treatments, induced a marked transcriptional up-regulation of key players of the recognition and repair of DNA double-strand breaks (Fig. 3C). Additionally, cell viability was reduced in combination with PNKP1 inhibition for TH10785, but not for TH5487 (Fig. 3, D and E). These results demonstrate that activation of the OGG1 β,δ-lyase activity by TH10785 occurs both in vitro and in cells and that PNKP1 is essential to avoid an accumulation of DNA damage and consequential cell death.

We present a new concept that describes the induction of an OGG1 β,δ-lyase activity by an enzyme-directed small-molecule catalyst that binds into the active site of the enzyme (Fig. 3F and figs. S27 and S28). The new catalytic function caused by the presence of TH10785 prefers AP sites over 8-oxoG and creates a dependency of PNKP1 in vitro and in cells. The improvement or rerouting of repair pathways that deal with oxidative DNA damage has implications in many conditions such as inflammation, cancer, Alzheimer's disease, and aging, and the concept outlined here allows for the possibility of controlling and rerouting repair pathways in a new manner (*24*).

REFERENCES AND NOTES

1. N. Donley *et al*., *ACS Chem. Biol.* **10**, 2334–2343 (2015).
2. Y. K. Tahara *et al*., *J. Am. Chem. Soc.* **140**, 2105–2114 (2018).
3. T. Visnes *et al*., *Science* **362**, 834–839 (2018).
4. B. A. Baptiste *et al*., *Free Radic. Biol. Med.* **124**, 149–162 (2018).
5. D. Schniertshauer *et al*., *DNA Repair* **87**, 102784 (2020).
6. S. Oka *et al*., *Sci. Rep.* **11**, 5819 (2021).
7. S. S. B. Komakula *et al*., *Sci. Rep.* **8**, 14886 (2018).
8. S. S. B. Komakula, B. Blaze, H. Ye, A. Dobrzyn, H. Sampath, *Int. J. Mol. Sci.* **22**, 1152 (2021).
9. W. You *et al*., *Angew. Chem. Int. Ed.* **56**, 1007–1011 (2017).
10. B. P. Hubbard *et al*., *Science* **339**, 1216–1219 (2013).
11. S. J. Gardell *et al*., *Nat. Commun.* **10**, 3241 (2019).
12. Z. Huang *et al*., *Nat. Chem. Biol.* **14**, 1118–1126 (2018).
13. J. C. Fromme, S. D. Bruner, W. Yang, M. Karplus, G. L. Verdine, *Nat. Struct. Biol.* **10**, 204–211 (2003).
14. M. Michel *et al*., *ACS Omega* **4**, 11642–11656 (2019).
15. T. Visnes *et al*., *DNA Repair* **71**, 118–126 (2018).
16. T. A. Rosenquist, D. O. Zharkov, A. P. Grollman, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7429–7434 (1997).
17. B. Baños, J. M. Lázaro, L. Villar, M. Salas, M. de Vega, *Nucleic Acids Res.* **36**, 5736–5749 (2008).
18. I. Boldogh *et al*., *J. Biol. Chem.* **287**, 20769–20773 (2012).
19. R. Jafari *et al*., *Nat. Protoc.* **9**, 2100–2122 (2014).
20. J. M. Baquero *et al*., *Sci. Rep.* **11**, 3490 (2021).
21. S. Roychoudhury *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11409–11420 (2020).
22. K. Kubo, H. Ide, S. S. Wallace, Y. W. Kow, *Biochemistry* **31**, 3703–3708 (1992).
23. P. Srivastava, A. Sarma, C. M. Chaturvedi, *PLOS ONE* **13**, e0190516 (2018).
24. N. Owen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016833118 (2021).

ACKNOWLEDGMENTS

We thank J. E. Unterlass, M. Kullman Magnusson, L. Sjöholm, T. Pham, S. Demir, A. Pliakou, T. Sandvall, K. Edfeldt, and M. Sundström for support and discussions. We thank the scientists at stations I04 and I04-1 of the Diamond Light Source, Didcot, Oxfordshire (UK), for their support during data collection (allocations MX15806 and MX21625). **Funding:** This work was supported by European Research Council grant TAROX-695376 (T.H.); Swedish Research Council grant 2015-00162 (T.H.); Swedish Research Council grant 2018-03406 (P.S.); Ministry of Science and Innovation, Spain/State Research Agency, Spain/10.13039/501100011033 and "ERDF A way to make Europe" grant BFU2017-83900-P (M.d.V.); Crafoord Foundation grant 20190532 (P.S.); an Alfred Österlund Foundation grant (P.S.); a Swedish Pain Relief Foundation grant (T.H.); Swedish Cancer Society grant CAN2018/0658 (T.H.); Swedish Cancer Society grant CAN 2017/716 (P.S.); a Torsten and Ragnar Söderberg foundation grant (T.H.); Dr. Åke-Olsson Foundation for Hematological Research grant 2020-00306 (M.M.); the Thomas Helleday Foundation for medical research postdoctoral stipends (M.M. and C.B.-B.); the NTNU Enabling Technology Programme on Biotechnology to Valentyn Oksenych (S.C.-Z.); EMBO Short-Term Fellowship 9005 (S.C.-Z.); a FEBS Short-Term Fellowship (S.C.-Z.); SSI Grants for Scandinavian Exchange (S.C.-Z.); German Research Foundation (DFG) 239748522 (A.J.K. and C.H.); Sonderforschungsbereich (SFB) 1127 (A.J.K. and C.H.); The Leibniz Award (C.H.); Norwegian Research Council grant 303369 (T.V.); Karolinska Institutet Research Foundation grant 2020-02186 (M.M.); a Lars Hiertas Minne Stiftelse grant (M.M.); and Asociacion Española Contra Cancer grant Postdoctoral AECC 2020, Nº POSTD20042BENI (C.B.-B.). This study was funded by "Instituto de Salud Carlos III CP19/00063 and PI20/00329" and cofunded by European Social Fund "Investigating in your future" and "European Regional Development Fund" (A.S.-P.). CNIO studies are partially funded by Instituto de Salud Carlos III, project reference PI19/00640, and cofounded by the European Regional Development Fund (ERDF), "A way to make Europe" (A.O.). This project received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement no. 875510 (M.M., E.J.H., E.W., and A.St.). This work received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 722729 (T.H. and B.M.F.H.). **Author contributions:** Conceptualization: M.M., C.B.-B., T.H.; Data curation: M.M., C.B.-B., B.M.F.H., O.M., G.M., J.D., O.W., K.S., S.C.-Z., A.-S.J., T.V., A.S.-P., J.J.A., E.J.H., Z.Z., C.N.C., A.Sa., A.J.K.; Formal analysis: M.M., C.B.-B., P.A.C., B.M.F.H., G.M., J.D., O.W., K.S., S.C.-Z., A.-S.J., T.V., A.S.-P., J.J.A., E.J.H., Z.Z., C.N.C., A.Sa., A.J.K., D.M., A.d.O., L.A., O.C.; Methodology: M.M., C.B.-B.; Resources: M.M., T.V., C.N.C., A.Sa., M.S., C.H., I.V., C.K., A.St., P.S., U.W.B., M.d.V., T.H.; Investigation: M.M., C.B.-B., P.A.C., B.M.F.H., O.M., G.M., J.D., O.W., K.S., S.C.-Z., A.-S.J., T.V., A.S.-P., J.J.A., A.N.D., E.J.H., K.Mar., Z.Z., C.N.C., A.Sa., E.W., C.v.N., A.J.K., V.R., S.M., E.C.H., M.V., E.R.S., M.P., S.K., H.H.-H., S.L., P.M., A.R., K.Mam., F.O., F.S., O.L., J.S., N.D.-E., T.K., M.H., A.d.O., L.A., O.C.; Visualization: M.M., C.B.-B., B.M.F.H., O.M., G.M., M.d.V.; Funding acquisition: M.M., C.B.-B., S.C.-Z., T.V., C.H., I.V., C.K., A.O., R.P., A.St., P.S., U.W.B., M.d.V., T.H.; Project administration: M.M.; Supervision: M.M., C.B.-B., O.M., O.W., K.S., T.V., M.S., C.H., I.V., C.K., A.O., R.P., A.St., P.S., U.W.B., M.d.V., T.H.; Writing – original draft: M.M., C.B.-B., T.H.; Writing – review and editing: M.M., C.B.-B., T.H. **Competing interests:** T.V., O.W., T.K., and T.H. are listed as inventors on US patent no. WO2019166639A1, which covers OGG1 inhibitors. The patent is fully owned by a nonprofit public foundation, the Thomas Helleday Foundation for Medical Research, and T.H. and U.W.B. are members of the foundation board. C.K. and U.W.B. are employees of Oxcia AB, a company that holds a license to WO2019166639A1. O.M., O.W., K.S., A.-S.J., T.V., E.J.H., E.W., O.L., T.K., M.H., M.S., C.K., P.S., U.W.B., and T.H. are shareholders of Oxcia AB. All remaining authors declare that they have no competing interests. **Data and materials availability:** Transcriptome data have been deposited at the Gene Expression Omnibus (GEO) with the accession number GSE188779. The atomic coordinates and structure factors (codes 7AYY, 7AYZ, and 7AY0) have been deposited in the Protein Data Bank (www.wwpdb.org/). All other data needed to evaluate the conclusions in this paper are present in either main text or the supplementary materials.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf8980
Materials and Methods
Supplementary Text
Figs. S1 to S39
Tables S1 to S10
References (*25*–*29*)
MDAR Reproducibility Checklist
Movies S1 to S3

View/request a protocol for this paper from *Bio-protocol*.

Submitted 27 November 2020; resubmitted 23 December 2021
Accepted 14 May 2022
10.1126/science.abf8980

CRISPR

Structural basis for RNA-guided DNA cleavage by IscB-ωRNA and mechanistic comparison with Cas9

Gabriel Schuler†, Chunyi Hu†, Ailong Ke*

Class 2 CRISPR effectors Cas9 and Cas12 may have evolved from nucleases in IS200/IS605 transposons. IscB is about two-fifths the size of Cas9 but shares a similar domain organization. The associated ωRNA plays the combined role of CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA) to guide double-stranded DNA (dsDNA) cleavage. Here we report a 2.78-angstrom cryo–electron microscopy structure of IscB-ωRNA bound to a dsDNA target, revealing the architectural and mechanistic similarities between IscB and Cas9 ribonucleoproteins. Target-adjacent motif recognition, R-loop formation, and DNA cleavage mechanisms are explained at high resolution. ωRNA plays the equivalent function of REC domains in Cas9 and contacts the RNA-DNA heteroduplex. The IscB-specific PLMP domain is dispensable for RNA-guided DNA cleavage. The transition from ancestral IscB to Cas9 involved dwarfing the ωRNA and introducing protein domain replacements.

Increasing evidence points to the possibility that the core components of the CRISPR-Cas adaptive immune systems evolved from genes in mobile genetic elements. The class 2 CRISPR effectors Cas9 (type II) and Cas12 (type V) are believed to have independently evolved from an ancestral TnpB-like nuclease, which is still commonly found in insertion sequence (IS) elements today (*1*–*3*). Cas9 appears to have emerged from a distinct branch of IS elements within the IS200/IS605 superfamily harboring IscB (*2*). IscB and Cas9 share a common domain architecture at the sequence level (Fig. 1, A and B). Both contain an arginine-rich bridge helix and an HNH endonuclease domain inserted into a RuvC endonuclease domain (*1*, *2*). The bridge helix in Cas9 plays a crucial role in mediating complex ribonucleoprotein (RNP) formation with two noncoding RNAs, CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA) (*4*–*8*). HNH and RuvC endonucleases are used by Cas9 to cleave the target and nontarget DNA strands, respectively (*4*). During CRISPR interference, the DNA substrate is validated through R-loop formation, which involves DNA unwinding and RNA-DNA heteroduplex formation (*7*, *9*–*14*). IscB was found to assemble with a single large [>200 nucleotides (nt)] noncoding RNA encoded by the transposon, ωRNA (OMEGA: obligate mobile element guided activity) (*2*). Together IscB-ωRNA mediates RNA-guided DNA cleavage similar to the Cas9-crRNA-tracrRNA RNP (*2*). To avoid self-targeting and to reduce search space, Cas9 further specifies a protospacer adjacent motif (PAM) adjacent to the target site (*15*, *16*). This mechanism is conserved in IscB-ωRNA, and the equivalent target-adjacent motif (TAM) is recognized by a TAM interaction domain (TID) in IscB (*2*). All IscBs further encode a PLMP-motif containing domain at the N terminus (*2*). This domain is not found in Cas9 and has an undefined function (Fig. 1B).

To understand the RNA-guided DNA cleavage mechanism by the compact IscB-ωRNA RNP and its relationship with Cas9-crRNA-tracrRNA, we determined a 2.78-Å structure of the gut microbiome–derived *Ogeu*IscB-ωRNA RNP complex (*2*) bound to target DNA using cryo–electron microscopy (cryo-EM) (Fig. 1, figs. S1 to S3, and movie S1). Whereas the majority of the 496–amino acid IscB and 222-nt ωRNA could be unambiguously resolved, only a portion of the 60–base pair (bp) DNA target could be reliably modeled. These include 13 bp of the TAM-proximal double-stranded DNA (dsDNA), the entire 16-nt target strand (TS) single-stranded DNA (ssDNA), and the 2-nt nontarget strand (NTS) ssDNA in the R-loop region (Fig. 2A). The TAM-distal DNA is missing from the EM density owing to molecular motion rather than cleavage and dissociation, because phosphorothioate modifications have been introduced into the DNA backbone at the HNH and RuvC cleavage sites (fig. S1H) (*2*).

We found that the architectural organization, domain functionality, and nucleic acid binding mode are similar between IscB-ωRNA and Cas9 RNP. IscB-ωRNA adopts a similar two-lobed architecture, although its overall shape is much flatter, because several surface

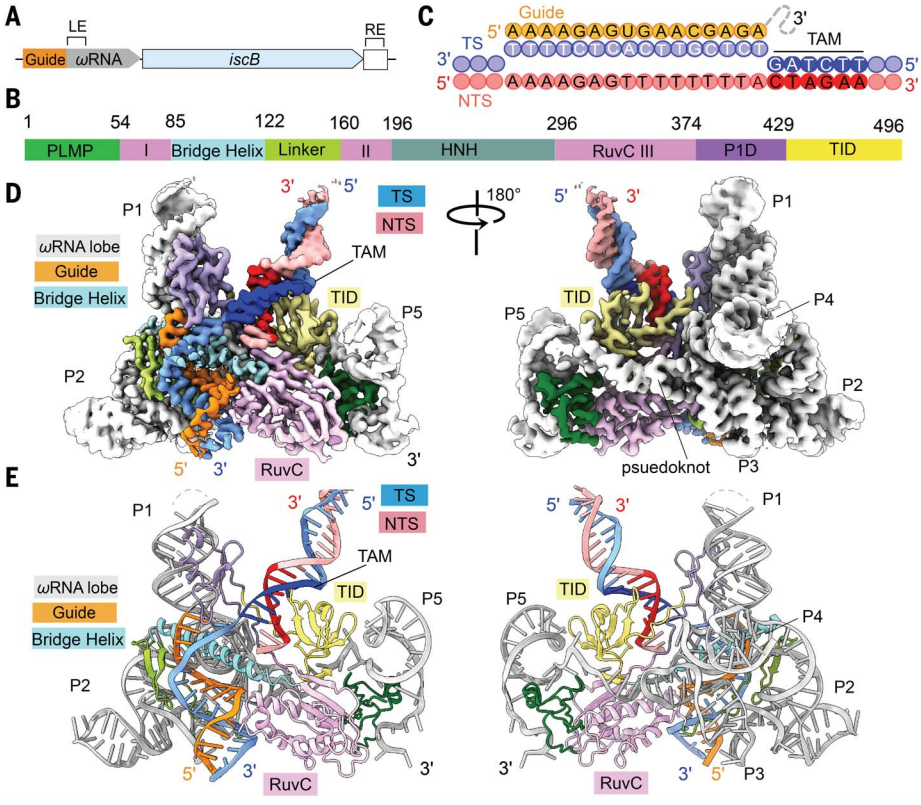


Fig. 1. Cryo-EM reconstruction and structure of IscB RNP bound to target DNA. (**A**) Arrangement of the *Ogeu*IscB and ωRNA in its native IS element defined by the left (LE) and right (RE) ends of the transposon. (**B**) Domain organization of IscB. P1D, P1 interaction domain; TID, TAM-interaction domain. RuvC domain is separated into three segments: RuvC I, II, and III. Color scheme is conserved throughout Fig. 1. (**C**) Diagram of R-loop formed between guide RNA and target DNA. TAM sequence is read 5′-CTAGAA-3′ on the nontarget strand. (**D** and **E**) Cryo-EM reconstruction at 2.78 Å and cartoon representations of the IscB-ωRNA/target DNA complex.

Department of Molecular Biology and Genetics, Cornell University, 253 Biotechnology Building, Ithaca, NY 14853, USA.
*Corresponding author. Email: ailong.ke@cornell.edu
†These authors contributed equally to this work.

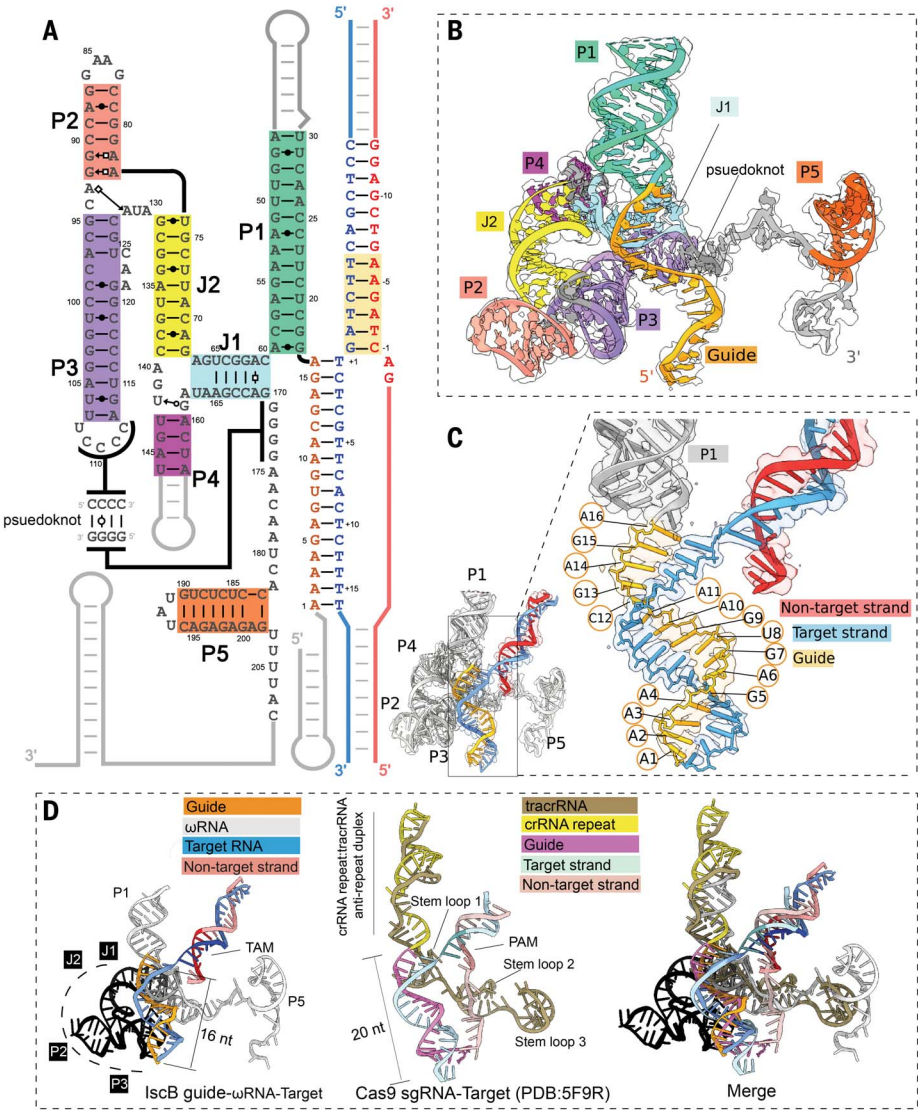


Fig. 2. Structural organization of the ωRNA and comparison to Cas9 crRNA-tracrRNA. (**A**) Schematic of ωRNA depicting secondary and tertiary interactions. Nontarget strand, red; target strand, blue; guide RNA, orange. (**B**) Atomic model of ωRNA. (**C**) Close-up view depicting R-loop base pairing between guide RNA and target strand DNA. (**D**) Structural alignment of ωRNA and tracrRNA-crRNA in *Sp*Cas9 RNP showing conserved RNA structures in guide RNAs, P1 with *Sp*Cas9 tracrRNA-crRNA helix, J1 with *Sp*Cas9 tracrRNA stem loop 1, P3 pseudoknot with *Sp*Cas9 tracrRNA stem loop 2, and P5 with *Sp*Cas9 tracrRNA stem loop 3. Colored in black is the region of the ωRNA replaced by the REC lobe in Cas9.

domains in Cas9 are missing in IscB (fig. S4 and movie S2). Structural alignments revealed that the P1 stem loop of ωRNA is the functional equivalent of the crRNA repeat–tracrRNA anti-repeat duplex in the Cas9 RNP. It occupies the same location in the RNP and assists R-loop formation in a similar manner, by stabilizing the guide-RNA/TS-DNA heteroduplex through continuous base stacking (Fig. 1, C to E, and movie S3). The TAM-containing dsDNA and the guide-RNA/TS-DNA heteroduplex in the R-loop region are accommodated by IscB-ωRNA at locations similar to those in Cas9s, through conceptually similar mechanisms (Fig. 1, D and E, and fig. S4). The TS-DNA base-pairs with the 16-nt guide RNA. The first 12 bp of the DNA-RNA heteroduplex adopts a distorted A-form owing to IscB contacts, with a widened major groove and base-stacking almost perpendicular to the helical axis. The last 4 bp of the heteroduplex adopts a canonical A-form geometry (Fig. 1, D and E).

Architecturally, the biggest structural difference between IscB and Cas9 is its lack of a polypeptide-based recognition (REC) lobe (fig. S4 and movie S2). The functional replacement is the ωRNA lobe (from J1 to the pseudoknot), which folds into a sophisticated tertiary RNA structure (Fig. 2A). The structured portion of ωRNA was previously identified as HEARO RNA [HNH endonuclease–associated RNA and ORF (open reading frame)] (*17*). This RNA and its associated HNH-containing ORF together were speculated to constitute a mobile genetic element (*17*). The three-dimensional (3D) structure supports the secondary structure models from previous studies (*2*, *17*). The central portion of ωRNA is a tail-to-tail stacked P2-P3 superhelix. J2 helix extrudes from the P2-P3 junction, then bifurcates into P4 and J1 at its end. Whereas P4 projects away, J1 projects toward the apex of P3. The following residues zip up with the apical loop of P3 through a 4-bp G/C-rich pseudoknot (Fig. 2, A and B). Following the pseudoknot, ωRNA extends horizontally along the back side of the IscB as a conserved single-stranded linker and a terminator-like element (P5, followed by four consecutive U's) (Fig. 2B). A conserved and highly structured RNA typically mediates either catalysis, ligand binding, or RNP formation (*17*). Our structure does not support a direct involvement of ωRNA in RNA-guided DNA cleavage because the bulk of ωRNA is insulated from the guide-RNA/TS-DNA heteroduplex by a layer of protein elements from IscB (Figs. 1, C and D, and 2C and movie S1). Our structure further suggests that the evolutionary trend from ancestral IscB to Cas9 involves replacing the structural roles of ωRNA with protein domains. However, the crRNA-tracrRNA of SpCas9 and NmeCas9 RNPs still contain structural elements reminiscent of P1, J1, pseudoknot, and terminator in ωRNA (Fig. 2D and fig. S5), presumably because these elements are indispensable for RNP assembly.

Opposite from the ωRNA lobe, the equivalent of the Cas9 nuclease (NUC) lobe contains the RuvC nuclease as its platform. RuvC is woven together from three split polypeptide elements (Fig. 1B, fig. S6A, and movie S1). It projects structural domains to various regions of the RNP. These elements are rich in positive surface charges, making favorable contacts with nucleic acids in different regions (fig. S6A). The N-terminal PLMP motif–containing domain is packed at the edge of the NUC lobe to capture the terminator-like structure in ωRNA (fig. S6B). The Arg-rich bridge helix is regarded as one of the most conserved structural elements in Cas9 (*7*, *8*). It plays an equally important function in IscB-ωRNA RNP. Projected from RuvC, the bridge helix travels underneath the guide RNA, along the pseudoknot and J1, and at the base of P1, making multiple electrostatic contacts to the sugar-phosphate backbones. A line of consecutive arginine and lysine residues along one phase of the bridge helix make consecutive phosphate contacts to seven residues in the RNA guide (U8-A14), immobilizing the seed region of the guide in place for TS-DNA base

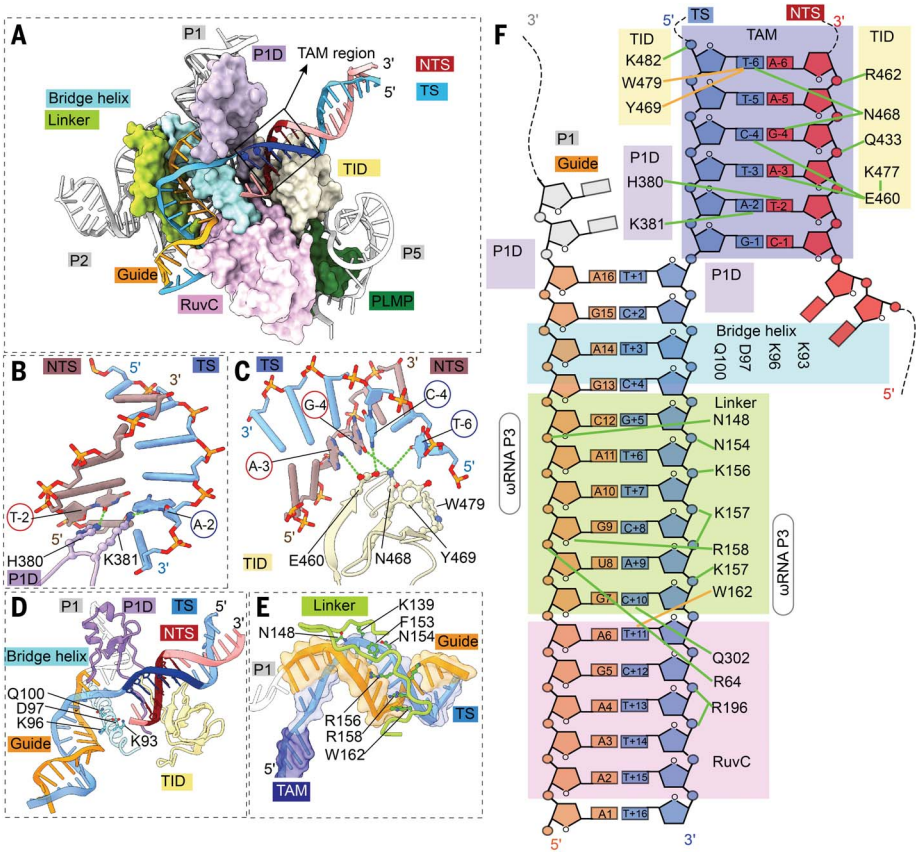


Fig. 3. Structural basis for TAM recognition and R-loop formation by IscB-ωRNA. (**A**) TAM recognition and R-loop specification by domains of IscB. Color scheme is consistent with that in Fig. 1. (**B**) Close-up view of P1 interaction domain (P1D) linker residues recognizing TAM-2 base pair (target adjacent motif) from the DNA minor groove side. (**C**) Close-up view of the IscB TAM interaction domain (TID) making base-specific contacts from the DNA major groove side. (**D**) Close-up view of the bridge helix and P1D making contacts with the beginning portion of the DNA-RNA heteroduplex in the R-loop region. (**E**) Close-up view of the β hairpin+linker domain specifying the minor groove of the middle portion of the DNA-RNA heteroduplex. (**F**) Diagram of IscB contacts to TAM and DNA-RNA heteroduplex in the R-loop. Positioning of the bridge helix domain separating the R-loop from the core of ωRNA in light blue. Green lines denote electrostatic contacts and brown lines denote hydrophobic contacts. TAM highlighted with purple box (ideal TAM sequence: 5′-NWRRNA-3′). guide RNA (orange), target strand DNA (blue), nontarget strand DNA (red).

pairing (fig. S6C). A β hairpin followed by a flexible linker connects the bridge helix back to RuvC. Although very degenerate in size and structural complexity, this flexible structural elements "glues" ωRNA and middle portion of the guide RNA together with its positive Arg and Lys residues (fig. S6D). The HNH nuclease domain is projected internally from RuvC. As in many Cas9s, this domain is not well resolved in the averaged EM density map owing to conformational flexibility. RuvC sends P1D domain to recognize the P1 helix of ωRNA; its functional equivalence is the WED domain in Cas9 (figs. S6E and S4) (*7*, *10*, *16*). Finally, P1D connects with the TAM-interaction domain (TID) situated above RuvC through flexible linkers.

The *Ogeu*IscB-ωRNA/R-loop structure explains the RNA-guided target recognition mechanism in high resolution (Fig. 3A and movie S4). TAM [5′-NWRRNA-3′ (*2*); actual sequence: CTAGAA] in the dsDNA target is captured from the major groove side by the TID domain of IscB and from the minor groove side by the P1D linker (Figs. 3B and 4, C and F). No contact was found at the −1 TAM position. The −2 TAM position is recognized from the minor groove side by His^{397} and Lys^{380} in P1D linker to O2 of T_{NTS-2} and N3 of A_{TS-2}, respectively. G-C pairs may be rejected in either combination owing to the steric clash caused by the N2 protrusion from guanosine into the minor groove. The −3 and −4 of TAM appear to be probed indirectly for shape complementarity. It is likely that only purines in the NTS support the van der Waals contacts to the backbone of Glu^{459} and Gly^{460} in TID; pyrimidines are too recessed. The −6 TAM position is recognized through hydrophobic contacts to the methyl groups of T_{TS-6} in the major grove, by Tyr^{468} and Trp^{478} in TID, respectively. Many IscB homologs encode smaller TID domains and specify less stringent TAM codes (*2*). Domain-swapping attempts, structure-guided design, and directed evolution could lead to more versatile IscB-ωRNA tools with expanded TAM codes.

A recent Cas9 study showed that off-targeting is inversely correlated with the extent of protein contacts to the guide-RNA/TS-DNA heteroduplex; the more local interactions to specify an A-form geometry, the less mismatch tolerance therein (*14*). In this regard, our structural analysis identified extensive R-loop contacts (Figs. 3A and 4, D to F), which implies that IscB-ωRNA can specify a DNA target stringently despite its miniature size and shorter R-loop specification. A P1D loop (amino acids 396 to 408) specifies the first two base pairs of the guide-RNA/TS-DNA heteroduplex from the minor groove side. The bridge helix and the following β hairpin and linker specify the middle portion of the heteroduplex (bp 2 to 9) from major and minor sides, respectively. ωRNA provides the platform support for these contacts, and a portion of the ωRNA backbone (P2, nt 114 to 116) directly contacts the backbone of guide RNA (bp 10 and 11). The RuvC domain then contacts the minor groove of bp 9 to 13. Base pairs 14 to 16 are not contacted and have weaker density. As shown later, this region is recognized when HNH docks onto the DNA-RNA heteroduplex.

To gain insight into the DNA cleavage mechanism, we analyzed the conformational dynamics in the IscB-ωRNA/R-loop EM reconstruction. Finer conformational sampling revealed two predominant conformational states. In the unlocked R-loop state (Fig. 4, A and B, and fig. S7), the 3.1-Å map shows the NTS-DNA traveling near the RNA-bound TS-DNA. NTS-DNA is blocked from accessing the RuvC active site because of a steric clash with the anchor connecting HNH to RuvC (Fig. 4A). Although unresolved in EM density, HNH is likely part of the blocking mechanism as well. Its approximate location can be inferred by comparing it to the NmeCas9 *apo* structure (*12*). By contrast, the 3.2-Å locked R-loop state (Fig. 4, C and D) shows HNH docking onto the RNA/TS-DNA heteroduplex and caging it with the rest of the IscB elements mentioned previously (Figs. 3A and 4C and movie S5). The entry and exiting linkers from RuvC to HNH probe for shape complementarity with the bottom and middle portions of the DNA-RNA heteroduplex, respectively. The body of HNH sinks into the major groove of the DNA-RNA heteroduplex (Fig. 4C). These close contacts are expected to further reduce mismatch tolerance. An Alphafold (*18*) predicted HNH

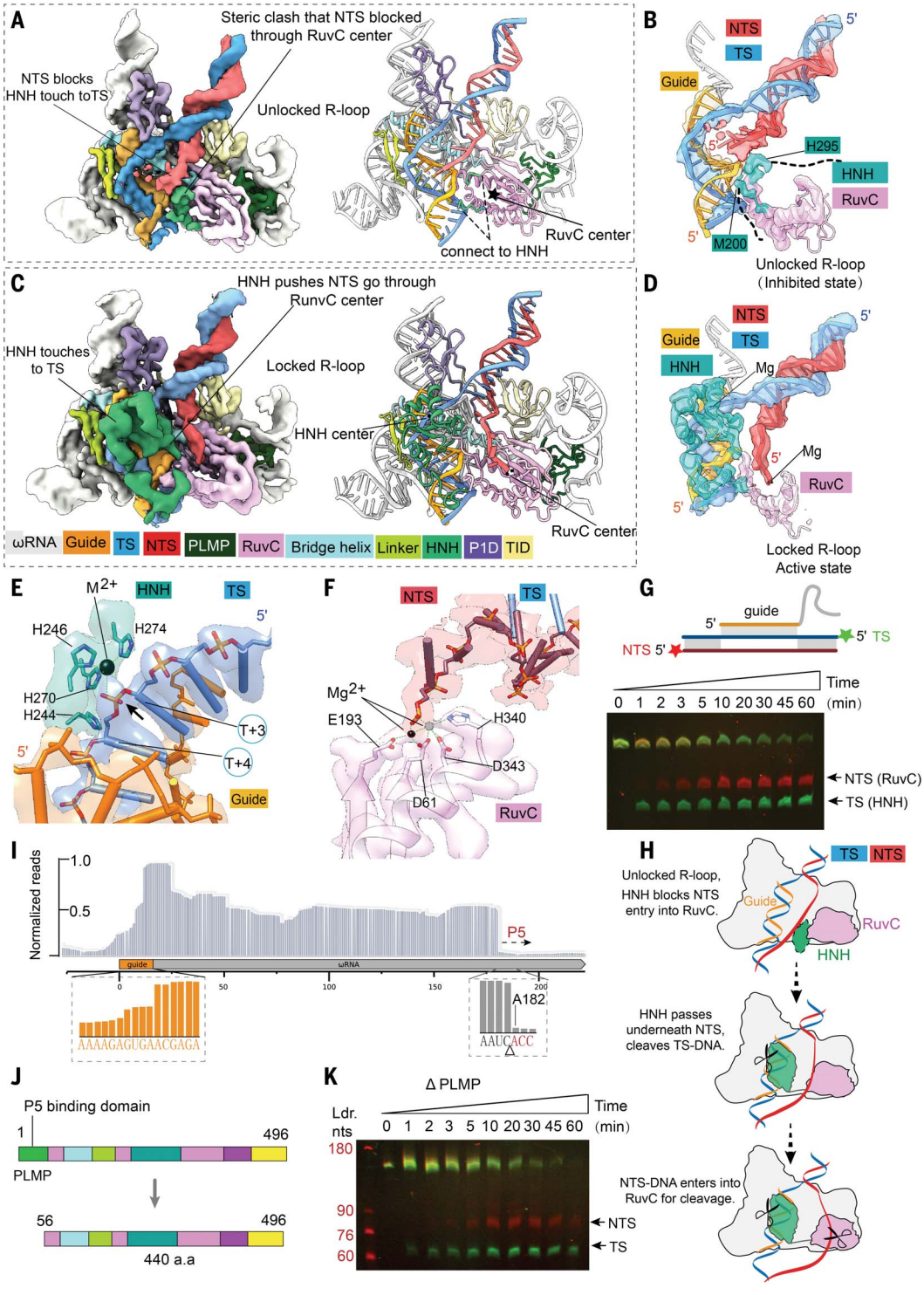


Fig. 4. Mechanistic dissection of RNA-guided DNA cleavage by IscB. (**A**) A 3.7-Å EM map and atomic model depicting the unlocked R-loop state. Color scheme is consistent with that in Fig. 1. (**B**) Focused view of DNA, guide RNA, and nuclease densities seen in the unlocked R-loop state. NTS is blocked from entering the RuvC cleavage site by the anchor of HNH to RuvC. (**C**) A 3.8-Å EM map and atomic model of the locked R-loop state. Alphafold predicted HNH domain structure (in green) is docked unambiguously into the EM density. Linker between HNH and RuvC domains can be seen interacting with the TAM-distal portion of the R-loop. (**D**) Focused view of HNH densities in the locked (active) state. The NTS density is now allowed into the RuvC active site. (**E**) Close-up view of the HNH active site in the locked state. Catalytic metal ion (black) is seen coordinated to the TS substrate. A second metal ion is required for cleavage (ball with dash line). It is repelled from the active site by the phosphorothioate modification in DNA. (**F**) Close-up view of the RuvC active site in the locked R-loop state. The coordinated catalytic metal ion (black) is seen contacting the backbone of the incoming NTS DNA. (**G**) Urea-PAGE (polyacrylamide gel electrophoresis) showing time-resolved DNA cleavage. TS is cleaved by HNH prior to NTS cleavage by RuvC, supporting the unlocked/locked R-loop cleavage model. (**H**) Proposed mechanistic model explaining ordered strand cleavage by IscB. (**I**) Small RNA sequencing of purified IscB-RNP, showing partial degradation of the guide RNA and a predictable cleavage site preceding stem-loop P5. (**J**) Domain organization of wild-type and ΔPLMP IscB. (**K**) Urea-PAGE showing time-resolved DNA cleavage by IscB ΔPLMP.8

structure was docked into EM map (Fig. 4, C and D, and fig. S8). Although the HNH core structure agreed with the density very well, manual adjustments were needed to fit the predicted linker structures into the density (fig. S8). The HNH nuclease "bites" onto the sugar-phosphate backbone of TS-DNA in the heteroduplex. The His-rich active site coordinates a catalytic metal ion toward the phosphate of the fourth residue in TS-DNA (Fig. 4E and fig. S9), which would leave 3 nt at the TS-DNA side after cleavage, consistent with the biochemistry (*2*). Topologically, the observed docking movement is only possible if HNH passes underneath NTS-DNA, in a game of limbo with NTS-DNA, which in turn clears the roadblock that previously denied NTS access to RuvC (movie S6). A continuous corridor of density reveals TAM-proximal NTS-DNA entering the RuvC active site, coordinated by a metal ion therein (Fig. 4F and figs. S10 and S11). The order of events explains the biochemical observation that TS-DNA cleavage precedes the NTS cleavage (Fig. 4, G and H). Previously, RuvC in SpCas9 was found to be allosterically controlled by HNH conformational changes (*19*), and its cleavage rate trails behind HNH

(*20*). Our structural analysis defines the structural basis for the allosteric control in IscB (Fig. 4H). The same mechanism is likely present in Cas9 RNP.

Given the robust RNA-guided deoxyribonuclease activity in vitro, it is puzzling to observe only weak genome editing activity from *Ogeu*IscB-ωRNA in human cells (*2*). We noticed the presence of multiple RNA species in the purified *Ogeu*IscB-ωRNA RNP and subjected the sample for RNA deep sequencing. The sequence coverage dropped immediately before the terminator-like P5 element of ωRNA (Fig. 4I). This is rather surprising because P5 density is clearly present in the RNP structure. We speculated that the cryo-EM particle picking and 3D reconstruction process might have inadvertently biased toward P5-containing single particles (fig. S2A). Given the high DNA cleavage activity in our *Ogeu*IscB-ωRNA RNP, we probed into the possibility that the PLMP-P5 interaction may be dispensable for RNA-guided DNA cleavage. Indeed, *Ogeu*IscB-ωRNA with a structure-guided PLMP domain truncation (Δaa1–55) was only slightly slower than the wild-type RNP in target DNA cleavage (Fig. 4, J and K, and fig. S12). This result argues that the PLMP domain is not ubiquitously essential for RNA-guided DNA cleavage among IscB homologs (*2*). We speculate that the PLMP-P5 interaction may instead be important for the biogenesis of IscB-ωRNA, by controlling the readthrough and termination ratio at ωRNA P5 to achieve copy-number balance between IscB and ωRNA. Alternatively, these domains may be important for the transposition of IS200/IS605. The sequencing result further revealed a stepwise decrease in coverage for the guide (after the 6th and 10th nucleotide; Fig. 4I). This pattern is consistent with the observed guide accessibility in the IscB-ωRNA structure (Fig. 1). Whether the guide RNA stability may have been the cause of low genome editing activity in IscB is worth investigation in the future. Indeed, naturally occurring tracrRNA variants containing an 11-nt-long guide were shown to convert *Sp*Cas9 from a nuclease to an RNA-guided transcriptional repressor (*21*). Chemical modification efforts also revealed that the guide RNA integrity could influence the in vivo activity of Cas9 substantially (*22*).

Our structural analysis provides a high-resolution explanation for the relationship between IscB-ωRNA and Cas9-crRNA-tracrRNA. ωRNA was speculated to transpose by itself on the basis of informatic searches (*17*); therefore, it may play more active roles than what the structures revealed. Such functions might no longer be needed, or even be detrimental, when IscB-ωRNA established co-option with the CRISPR system. This may have led to the observed adaptation in Cas9-crRNA-tracrRNA, where the body of ωRNA was entirely replaced by protein domains. The remaining portion serves only two essential functions: guiding RNP assembly and guiding target searching. On the application side, there has been a strong interest to miniaturize Cas9 for expanded usage. For example, it would be desirable to package the next-generation Cas9-based genome editors (*23*–*27*) into mature delivery tools, such as the adenovirus-associated virus (AAV) vectors. Neither structure-guided approach nor directed evolution was particularly successful at miniaturizing the RNA-guided nucleases. By peeking into nature's winning solutions, we gain a fresh starting point to develop a new generation of powerful genome editing tools, packageable into AAV. Fifty-five amino acids have already been removed from IscB without abolishing its activity (Fig. 4I). Further structure-guided efforts will likely lead to smaller, more robust, and more active versions of genome editors.

REFERENCES AND NOTES

1. V. V. Kapitonov, K. S. Makarova, E. V. Koonin, *J. Bacteriol.* **198**, 797–807 (2015).
2. H. Altae-Tran *et al.*, *Science* **374**, 57–65 (2021).
3. T. Karvelis *et al.*, *Nature* **599**, 692–696 (2021).
4. M. Jinek *et al.*, *Science* **337**, 816–821 (2012).
5. L. Cong *et al.*, *Science* **339**, 819–823 (2013).
6. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2579–E2586 (2012).
7. H. Nishimasu *et al.*, *Cell* **156**, 935–949 (2014).
8. S. Shmakov *et al.*, *Mol. Cell* **60**, 385–397 (2015).
9. F. Jiang *et al.*, *Science* **351**, 867–871 (2016).
10. M. Jinek *et al.*, *Science* **343**, 1247997 (2014).
11. H. Nishimasu *et al.*, *Cell* **162**, 1113–1126 (2015).
12. W. Sun *et al.*, *Mol. Cell* **76**, 938–952.e5 (2019).
13. A. Das *et al.*, *Nat. Commun.* **11**, 6346 (2020).
14. J. P. K. Bravo *et al.*, *Nature* **603**, 343–347 (2022).
15. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, *Microbiology (Reading)* **155**, 733–740 (2009).
16. C. Anders, O. Niewoehner, A. Duerst, M. Jinek, *Nature* **513**, 569–573 (2014).
17. Z. Weinberg, J. Perreault, M. M. Meyer, R. R. Breaker, *Nature* **462**, 656–659 (2009).
18. J. Jumper *et al.*, *Nature* **596**, 583–589 (2021).
19. S. H. Sternberg, B. LaFrance, M. Kaplan, J. A. Doudna, *Nature* **527**, 110–113 (2015).
20. S. Gong, H. H. Yu, K. A. Johnson, D. W. Taylor, *Cell Rep.* **22**, 359–371 (2018).
21. R. E. Workman *et al.*, *Cell* **184**, 675–688.e19 (2021).
22. A. Mir *et al.*, *Nat. Commun.* **9**, 2641 (2018).
23. N. M. Gaudelli *et al.*, *Nature* **551**, 464–471 (2017).
24. A. V. Anzalone *et al.*, *Nature* **576**, 149–157 (2019).
25. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, *Nature* **533**, 420–424 (2016).
26. A. V. Anzalone *et al.*, *Nat. Biotechnol.* **40**, 731–740 (2022).
27. E. I. Ioannidi *et al.*, bioRxiv 2021.2011.2001.466786 (2021).

ACKNOWLEDGMENTS

We thank F. Ding for carrying in vitro RNA transcriptions to test IscB RNP activities, Q. Sun for providing initial IscB Alphafold predictions, and A. Pyle for advice on cryo-EM grid preparation. **Funding:** National Institutes of Health grant GM118174 (A.K.), Department of Defense through the National Defense Science & Engineering Graduate Fellowship Program (G.S.), National Science Foundation Materials Research Science and Engineering Centers program grant DMR-1719875 [Cornell Center for Materials Research Shared DOE Office of Biological and Environmental Research grant KP1607011 (Laboratory for BioMolecular Structure)]. **Author contributions:** Conceptualization: A.K., G.S., C.H.; Methodology: G.S., C.H., A.K. Investigation: G.S., C.H., A.K. Visualization: C.H., G.S., A.K. Funding acquisition: A.K., G.S. Project administration: A.K. Supervision: A.K. Writing – original draft: A.K., G.S., C.H. Writing – review and editing: A.K., G.S., C.H. **Competing interests:** The authors declare competing financial interests on IscB-ωRNA structure-inspired genome editing applications. **Data and materials availability:** The IscB-ωRNA/R-loop coordinates and cryo-EM density map have been deposited in the Protein Data Bank (PDB:7UTN) and the Electron Microscopy Data Bank (EMD-26782). The RNA-sequencing data has been deposited in the Sequence Read Archive (PRJNA838366). Plasmids used in this study are available upon request.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq7220
Materials and Methods
Figs. S1 to S12
Table S1
References (*28*)
MDAR Reproducibility Checklist
Movies S1 to S6

View/request a protocol for this paper from *Bio-protocol*.

Submitted 29 April 2022; accepted 18 May 2022
Published online 26 May 2022
10.1126/science.abq7220

ORGANIC CHEMISTRY

Catalyst-controlled site-selective methylene C–H lactonization of dicarboxylic acids

Hau Sun Sam Chan[1]†, Ji-Min Yang[1]†, Jin-Quan Yu[1]*

Catalyst-controlled site-selective activation of β- and γ-methylene carbon-hydrogen (C–H) bonds of free carboxylic acids is a long-standing challenge. Here we show that, with a pair of palladium catalysts assembled with quinoline-pyridone ligands of different chelate ring sizes, it is possible to perform highly site-selective monolactonization reactions with a wide range of dicarboxylic acids, generating structurally diverse and synthetically useful γ- and δ-lactones via site-selective β- or γ-methylene C–H activation. The remaining carboxyl group serves as a versatile linchpin for further synthetic applications, as demonstrated by the total synthesis of two natural products, myrotheciumone A and pedicellosine, from abundant dicarboxylic acids.

Achieving site-selective methylene C–H activation of aliphatic carboxylic acids is a long-standing challenge in organic synthesis. The freedom to site-selectively functionalize particular methylene C–H bonds in these ubiquitous molecular backbones would enable rapid construction of a great variety of molecular scaffolds and usher in retrosynthetic pathways that were previously deemed implausible. Although directed C–H activation of aliphatic carboxylic acids has been achieved with platinum (*1*, *2*) and palladium catalysis (*3*–*11*), development of catalysts for the activation of methylene C–H bonds is nascent, and identification of methods for distinguishing between the similar C–H bonds of multiple adjacent methylene units is an enduring challenge. Recent breakthroughs in ligand development have provided a glimmer of hope for the palladium-catalyzed, carboxylic acid–directed activation of β-methylene C–H bonds, culminating in reports of a dehydrogenation reaction and a deuteration reaction (*12*, *13*). However, no other C–heteroatom bond-formation reaction or activation of the γ-methylene C–H bonds has yet been identified with this approach. Although alternative strategies exist for the functionalization of methylene C–H bonds of free carboxylic acids via high-valent, electrophilic Fe or Mn catalysts (*14*, *15*), these methods display limited differentiation between methine and methylene C–H bonds. Hence, the challenge of highly methylene-selective, catalyst-controlled site-selective C–H functionalization between adjacent methylene units remains unmet.

Because lactones are ubiquitous in organic chemistry and monolactonization of dicarboxylic acids would leave the remaining carboxyl groups as versatile handles for further synthetic elaborations (*16*), we aimed to develop site-selective intramolecular C–O bond-formation reactions of dicarboxylic acids through lactonization. It is estimated that more than 3000 γ-lactones exist in nature (*17*, *18*), and the formation of lactones at methylene carbons is commonplace in natural products, as demonstrated in the structures and total syntheses of ginkgolide B (*19*, *20*), bilobalide (*21*, *22*), picrotoxinin (*23*–*25*), and cladantholide (*26*). Benzo-fused lactones such as phthalides and isochromanones are also recurring structures in bioactive natural products such as colletotrialide, fusarentin 6,7-dimethyl ether, and ajudazol B (*27*, *28*) (Fig. 1A). The principal challenge for the development of the desired lactonization reactions was to achieve tunable site-selective activation of either β- or γ-methylene C–H bonds by catalyst design. Another challenge was to achieve regiocontrol for unsymmetrical dicarboxylic acids, thereby precluding the unselective generation of mixtures of regioisomeric products (*29*) (Fig. 1B).

Here we report two distinct catalysts for site-selective β- and γ-C–H activation of dicarboxylic acids using quinoline-pyridone ligands, leading to three modes of lactonization for the synthesis of γ- and δ-lactones from various dicarboxylic acids (Fig. 1C). Symmetrical dicarboxylic acids, such as adipic acid and pimelic acid, could be converted into unsymmetrical lactone acids by means of this methodology. Unsymmetrical dicarboxylic acids generally produced lactone acids in which the more substituted fraction of the molecule preferentially lactonized. These lactonization reactions convert dicarboxylic acids into trifunctional molecules with three orthogonal sites of reactivity that are amenable to chemoselective transformations for complex molecule synthesis. We also discovered that the use of cheap and abundant MnO_2 as the oxidant is compatible with the lactonization reactions, demonstrating the possibility of using silver-free conditions for palladium-catalyzed, carboxylic acid–directed lactonization of methylene C–H bonds. The utility of this methodology was demonstrated with two total syntheses, forging the natural products pedicellosine and myrotheciumone A from dicarboxylic acids.

The development of the lactonization reaction began with exploratory studies in which adipic acid (**1**) was used as the standard substrate. We were mindful that bis-carboxylate chelation by the dicarboxylic acid could prevent the desired directed C–H activation, and thus a tightly bound bidentate ligand would be required to overcome this potential hurdle. Building on the precedent that pyridine-pyridone–type ligands could promote β-methylene C–H activation in a palladium-catalyzed, carboxylic acid–directed dehydrogenation reaction (*13*), a series of pyridine-pyridone–type ligands were investigated for the desired C–O bond-forming reactivity (tables S1 to S11). After extensive optimization of reaction parameters, the five-membered chelating quinoline-pyridone ligand **L1** emerged as the most effective at initiating the lactonization of **1** to provide γ-lactone **1a** in 65% isolated yield (Fig. 2A, first product), with *p*-xyloquinone as an additive. The rationale for the use of *p*-xyloquinone comes from the literature precedent that it may assist reductive elimination at the palladium center (*30*). The formation of **1a** from **1** implied two mechanistic possibilities: γ-lactonization via γ-C–H activation or γ-lactonization via β-C–H activation. In other words, the observed γ-lactonization could occur by means of directed γ-C–H activation followed by γ-lactonization by the same carboxyl group, or it could occur through directed β-C–H activation by one of the carboxyl groups followed by γ-lactonization by the remaining carboxyl group. To distinguish between these two possibilities, a control experiment using pimelic acid (**2**) as the substrate was carried out. We found that δ-lactone **2a** was generated in this experiment in 25% isolated yield (Fig. 2B, first product). This observation suggests that carboxylic acid–directed β-C–H activation with ligand **L1** is more probable, which is then followed by δ-lactonization using the carboxylic acid at the other end of the molecule. The other possibility for δ-lactonization, via a δ-C–H activation pathway, was deemed highly unlikely. Consistent with this hypothesis, the monomethyl ester of **1**, monomethyl adipate (**3**), was found to be unreactive for γ-lactonization. In the absence of the second carboxyl group in **3**, dehydrogenation of **3** to generate the corresponding α,β-unsaturated acid **4** was observed in ~17% nuclear magnetic resonance (NMR) yield under the γ-lactonization conditions (fig. S2). However, use of (*E*)-hex-2-enedioic acid (**5**)—the α,β-dehydrogenation product of **1**—as the substrate provided substantially reduced yield of γ-lactone **1a** (24%

[1]Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.
†These authors contributed equally to this work.
*Corresponding author. Email: yu200@scripps.edu

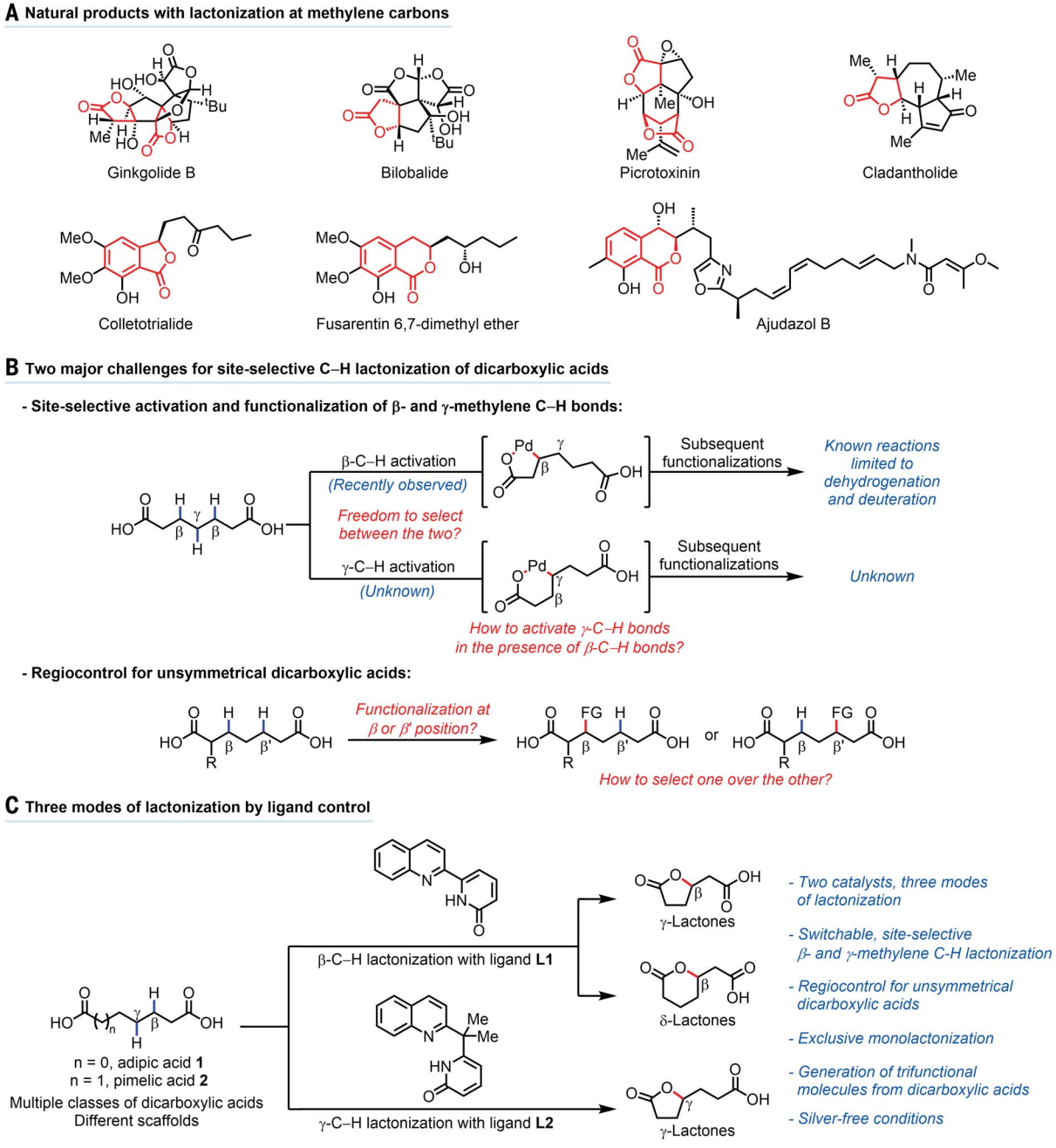

A Natural products with lactonization at methylene carbons
Ginkgolide B
Bilobalide
Picrotoxinin
Cladantholide
Colletotrialide
Fusarentin 6,7-dimethyl ether
Ajudazol B
B Two major challenges for site-selective C–H lactonization of dicarboxylic acids
- Site-selective activation and functionalization of β- and γ-methylene C–H bonds:
β-C–H activation
(Recently observed)
Freedom to select between the two?
γ-C–H activation
(Unknown)
Subsequent functionalizations
Known reactions limited to dehydrogenation and deuteration
Unknown
How to activate γ-C–H bonds in the presence of β-C–H bonds?
- Regiocontrol for unsymmetrical dicarboxylic acids:
Functionalization at β or β' position?
or
How to select one over the other?
C Three modes of lactonization by ligand control
β-C–H lactonization with ligand L1
γ-C–H lactonization with ligand L2
γ-Lactones
δ-Lactones
γ-Lactones
n = 0, adipic acid 1
n = 1, pimelic acid 2
Multiple classes of dicarboxylic acids
Different scaffolds
- Two catalysts, three modes of lactonization
- Switchable, site-selective β- and γ-methylene C-H lactonization
- Regiocontrol for unsymmetrical dicarboxylic acids
- Exclusive monolactonization
- Generation of trifunctional molecules from dicarboxylic acids
- Silver-free conditions

Fig. 1. Lactonization via activation of methylene C–H bonds. (**A**) Natural products with lactonization at methylene carbons. (**B**) Two major challenges for site-selective C–H lactonization of dicarboxylic acids. R, generic substituent. (**C**) Three modes of lactonization by ligand control.

NMR yield) under the γ-lactonization conditions (fig. S3). These observations suggest that one carboxyl group is responsible for directed β-C–H activation and the other is responsible for γ-lactonization, and that this lactonization process is unlikely to occur solely via a dehydrogenation–intramolecular lactonization sequence.

The generality of these two modes of lactonization via β-C–H activation was investigated with a series of dicarboxylic acids with various substituents and topologies. For γ-lactonization reactions via β-C–H activation with ligand **L1** (Fig. 2A), substrates possessing α-hydrogens that generally correlate with less-efficient C–H activation all provided the desired products in synthetically useful yields (*5*). The α-mono-substituted products with increasing steric demand of the substituent [methyl (Me), **6a**; ethyl (Et), **7a**; isopropyl (^{i}Pr), **8a**; benzyl (Bn), **9a**; *tert*-butyl (^{t}Bu), **10a**] were isolated in 50 to 60% yield. However, the product **11a**, which has a phthalimidoyl (NPhth) substituent at its α position, provided only a 25% isolated yield of the benzyl ester. The low yield of **11a** was due to low conversion, as the unreacted starting material was the major species observed on crude ^{1}H NMR. For the product **12a**, which bears the *gem*-dimethyl substituent at its α position, the carboxylic acid was isolated in 84% yield. Increasing the substituent size to a *gem*-diethyl system, as for product **13a**, resulted in a decrease in isolated yield to 66%. Products **14a** and **15a**, which have substitution at their β position but fully unsubstituted α positions, displayed high

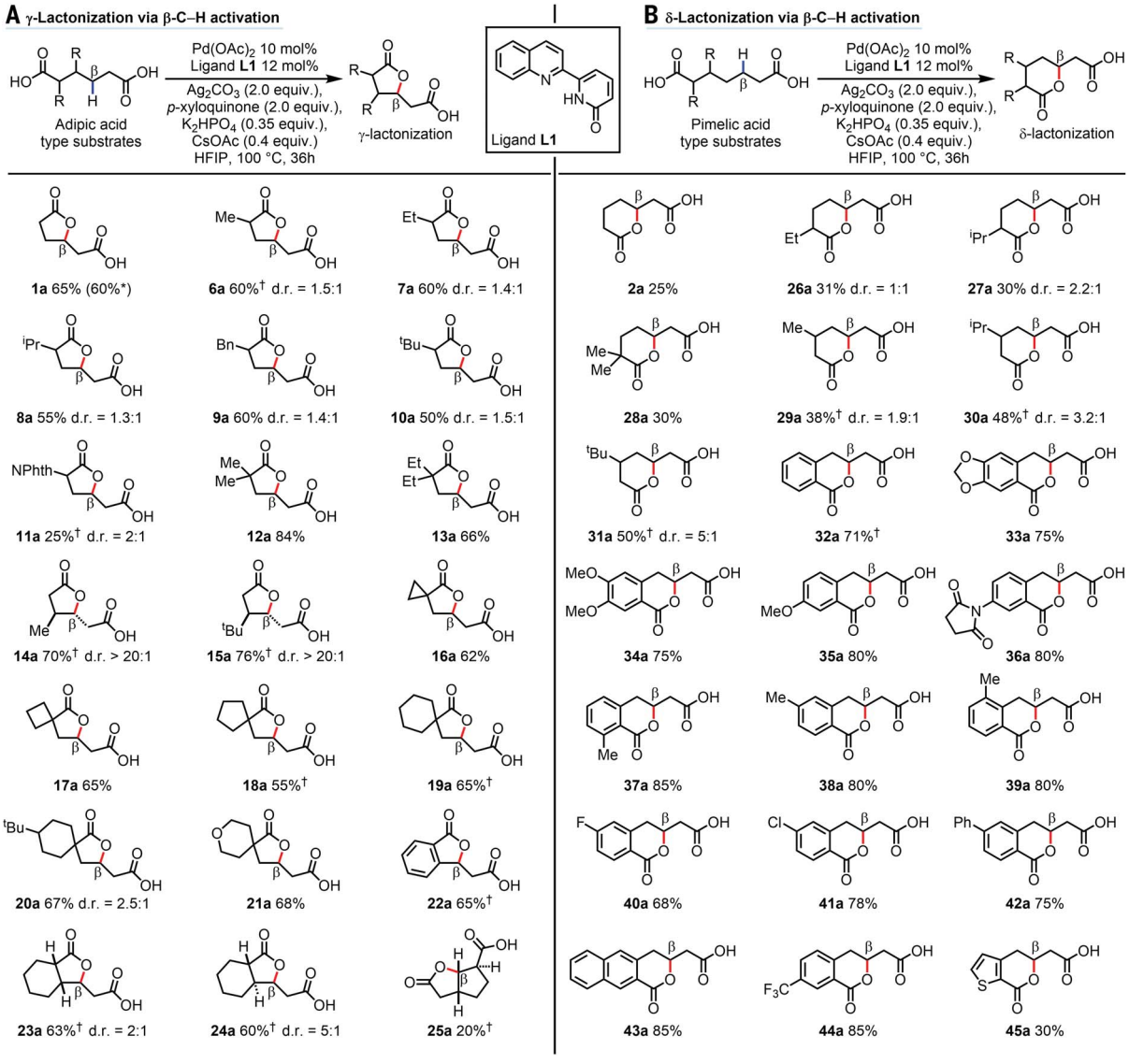


Fig. 2. Substrate scope of γ- and δ-lactones from β-C–H activation. (**A**) γ-Lactonization via β-C–H activation. (**B**) δ-Lactonization via β-C–H activation. Reaction conditions: substrate (0.1 mmol), $Pd(OAc)_2$ 10 mol %, ligand **L1** 12 mol %, Ag_2CO_3 (2.0 equiv), *p*-xyloquinone (2.0 equiv), K_2HPO_4 (0.35 equiv), CsOAc (0.4 equiv), 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP) (1.0 ml), 100°C, 36 hours. Isolated yields are reported. *Reaction conducted with 1.0 g of adipic acid (**1**). †Isolated as the corresponding benzyl ester. OAc, acetoxy; NPhth, phthalimidoyl; d.r., diastereomeric ratio.

reactivity and diastereoselectivity, generating isolated yields around 70% as their benzyl esters and favoring the anti isomer with a diastereomeric ratio >20:1. Spirocyclic structures at the α position provided lactones that have spirocyclic cyclopropyl (**16a**), cyclobutyl (**17a**), cyclopentyl (**18a**), cyclohexyl (**19a** and **20a**), and 4-tetrahydropyranyl (**21a**) systems in 55 to 68% yield. Fused structures were also found to be well tolerated for this lactonization methodology, providing benzo-fused product **22a** in 65% isolated yield as its benzyl ester, as well as fused 5,6-bicyclic systems with a cis (**23a**) and a trans (**24a**) junction in respective yields of 63 and 60% as their benzyl esters. The more strained 5,5-bicyclic system **25a** with endocyclic β-methylene C–H bonds was isolated in just 20% yield as its benzyl ester; the low yield was again due to low conversion. We also investigated the scalability of this γ-lactonization reaction via β-C–H activation, in which the lactone **1a** could be obtained in 60% isolated yield with 1.0 g of **1** as starting material (see supplementary materials for details on experimental setup).

For δ-lactonization reactions via β-C–H activation (Fig. 2B), aliphatic systems were challenging substrates (**2a** and **26a** to **31a**), and substitution at the β position was required to elevate the isolated yield to 38 to 50% (**29a** to **31a**). However, the reaction performance was superior for the synthesis of benzo-fused products, providing **32a** to **45a** in 68 to 85% yield and tolerating a range of aromatic substituents such as oxygen (**33a** to **35a**), nitrogen (**36a**), alkyl (**37a** to **39a** and **42a**), halide (**40a** and **41a**), and a trifluoromethyl group (**44a**). Modifying the identity of the aromatic ring was also possible, as demonstrated by the formation of naphthalene-fused system **43a** and thiophene-fused system **45a**, though the latter was isolated in only 30% yield.

Next, we wondered whether the γ-methylene C–H bonds of **2** could also be activated for γ-lactonization. This selectivity presented a particular challenge for a C–H activation approach, because activation of the competing β-methylene C–H bonds that would result in five-membered cyclopalladation would tend to be both kinetically and thermodynamically

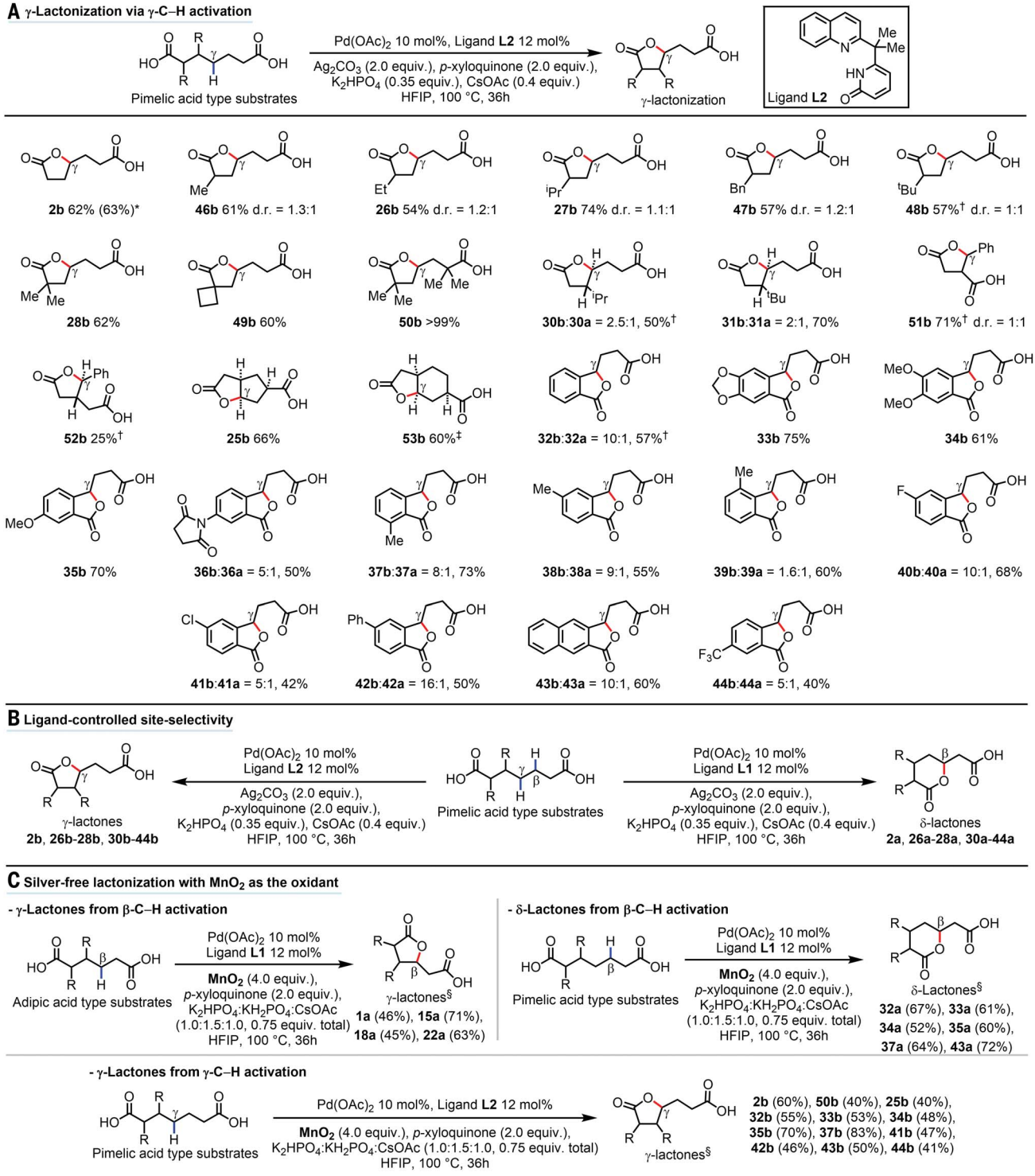


Fig. 3. Substrate scope of γ-lactones from γ-C–H activation, demonstration of switchable site selectivity, and examples of silver-free lactonization reactions with MnO_2 as the oxidant. (**A**) γ-Lactonization via γ-C–H activation. Reaction conditions: substrate (0.1 mmol), $Pd(OAc)_2$ 10 mol %, ligand **L2** 12 mol %, Ag_2CO_3 (2.0 equiv), *p*-xyloquinone (2.0 equiv), K_2HPO_4 (0.35 equiv), CsOAc (0.4 equiv), HFIP (1.0 ml), 100°C, 36 hours. Isolated yields are reported. *Reaction conducted with 1.0 g of pimelic acid (**2**). †Isolated as the corresponding benzyl ester. ‡Based on reactive diastereomer (see supplementary materials pp. S92 and S93 for explanation). (**B**) Demonstration of switchable site selectivity (19 examples). (**C**) Examples of silver-free lactonization with MnO_2 as the oxidant (22 examples). Reaction conditions: substrate (0.1 mmol), $Pd(OAc)_2$ 10 mol %, ligand **L1** or **L2** 12 mol %, MnO_2 (4.0 equiv), *p*-xyloquinone (2.0 equiv), K_2HPO_4:KH_2PO_4:CsOAc (1.0:1.5:1.0, 0.75 equiv in total), HFIP (1.0 ml), 100°C, 36 hours. §NMR yields.

preferred (*31*). Overcoming this hurdle for C–H activation could lead to two distinct site-selective lactonization pathways via β- or γ-methylene C–H activation. With this objective in mind, we further explored other types of quinoline-pyridone ligands and discovered that ligand **L2** provided the desired γ-lactone **2b** in 62% isolated yield (Fig. 3A). Evidence to support a γ-C–H activation process was obtained from H/D exchange and a series of control experiments (figs. S1 and S4). For this mode of lactonization via γ-C–H activation, substrates with α-hydrogens that generally correlate with less-efficient C–H activation all provided the desired products in synthetically useful yields (*5*). The α-mono-substituted products with increasing steric demand of the substituent (Me, **46b**; Et, **26b**; ^{i}Pr, **27b**; Bn, **47b**; ^{t}Bu, **48b**) were isolated in 54 to 74% yield. For the products that are quaternized at their α positions, the product **28b** (with single *gem*-dimethyl substitution) was isolated in 62% yield. For the product **50b** (with double *gem*-dimethyl substitutions), the isolated yield increased to >99%. The formation of spirocyclic systems was also tested with this lactonization methodology; however, only the cyclobutyl system was amenable to lactonization, producing **49b** in an isolated yield of 60%. γ-Lactones **30b** and **31b** that have substitution at their β position but fully unsubstituted α positions were isolated as the anti-diastereomers, along with separable minor products (δ-lactones **30a** and **31a**) in an ~2:1 ratio, with a total yield of 50 to 70%. Dicarboxylic acids with structures that would allow lactonization to occur on branched substituents but not on the parent alkyl chain were also investigated. Notably, the product **51b**, arising from α-benzyl succinic acid, was produced with the use of ligand **L2** but not with ligand **L1**, which suggests that a γ-C–H activation pathway was likely for its lactonization. This product was isolated in 70% yield after conversion to its benzyl ester. The product **52b**, arising from β-benzyl glutaric acid, was isolated as its benzyl ester in 25% yield as a single anti-diastereomer. The low yield for **52b** was due to low conversion. Substrates with endocyclic γ-methylene C–H bonds were also found to be viable for this lactonization protocol, as exemplified with products **25b** and **53b**, generating lactones with fused 5,5-bicyclic and fused 5,6-bicyclic systems in good yields as single diastereomers. Benzo-fused γ-lactones (**32b** to **44b**) were also forged in 40 to 75% total yields with predominant (γ:β > 8:1) site selectivity at the γ position (*32*, *33*), except for products **36b**, **39b**, **41b**, and **44b**, in which the site selectivity was found to be lower. The collective synthesis of a series of γ-lactones (**2b**, **26b** to **28b**, and **30b** to **44b**) and δ-lactones (**2a**, **26a** to **28a**, and **30a** to **44a**) from the same substrates demonstrates the desired ligand-controlled switchable site-selectivity for carboxylic acid–directed β-methylene and γ-methylene C–H lactonizations (Fig. 3B). We also investigated the scalability of this γ-lactonization reaction via γ-C–H activation, in which the lactone **2b** could be obtained in 63% isolated yield with 1.0 g of **2** as starting material (see supplementary materials for details on experimental setup).

With the scope of the lactonization established, we aimed to identify silver-free conditions for the three modes of lactonization reactions and found that MnO_2 was a viable replacement for Ag_2CO_3 as the oxidant. Upon further fine-tuning of the reaction conditions (tables S2 and S11), the lactonization reactions could be rendered silver-free across a series of substrates (Fig. 3C).

In light of the broad scope of lactones prepared via this methodology, we turned our attention toward two total syntheses as a means to demonstrate the synthetic utility offered by the fusion of dicarboxylic acids and C–H lactonization (Fig. 4). Myrotheciumone A (**64**)

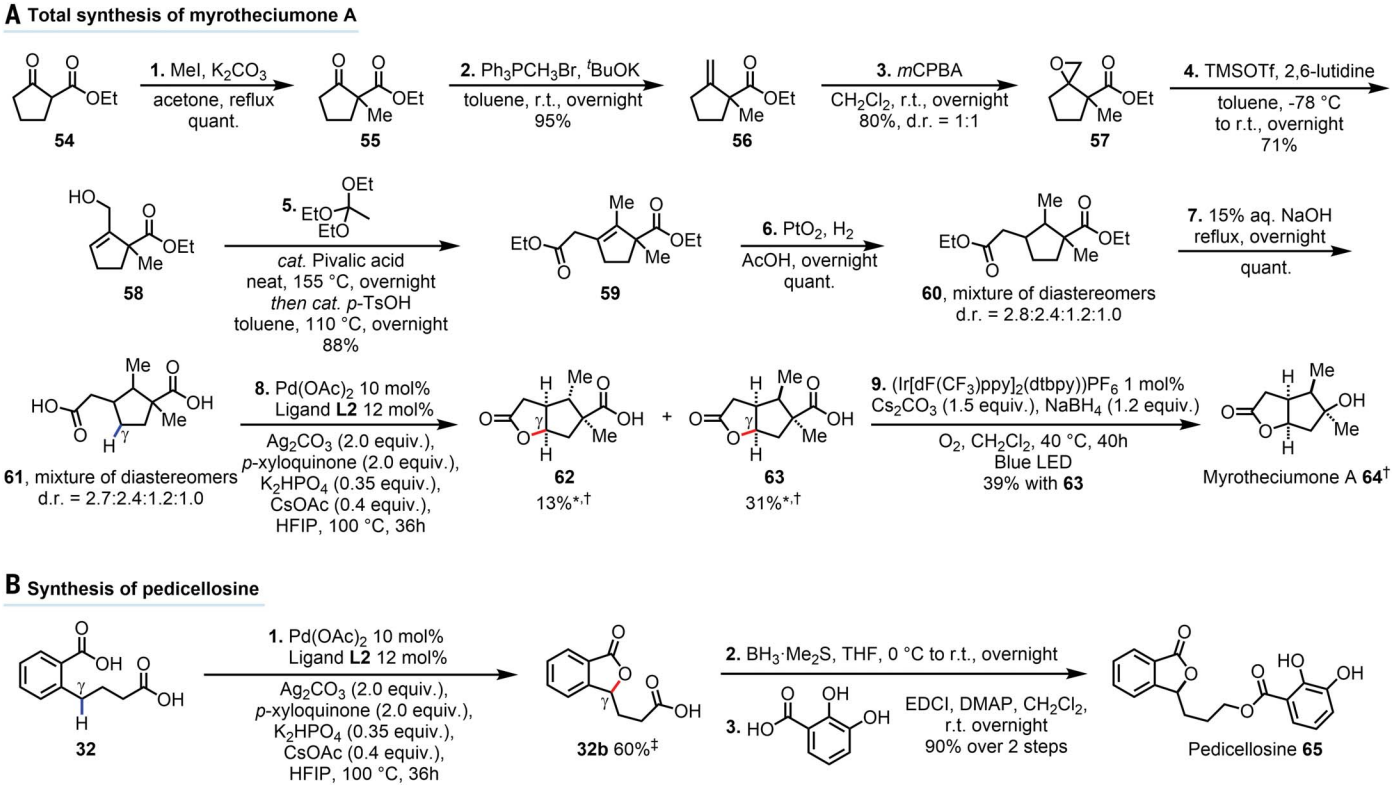


Fig. 4. Syntheses of natural products. (**A**) Total synthesis of myrotheciumone A. *Yields based on reactive diastereomers of **61** (see supplementary materials p. S118 for detailed analysis). †Relative configurations are shown, as the synthesis is racemic. (**B**) Synthesis of pedicellosine. ‡Isolated as the carboxylic acid. *m*CPBA, *meta*-chloroperoxybenzoic acid; TMSOTf, trimethylsilyl trifluoromethanesulfonate; EDCI, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide; DMAP, 4-dimethylaminopyridine; THF, tetrahydrofuran; dF(CF_3)ppy, 2-(2,4-difluorophenyl)-5-(trifluoromethyl) pyridine; dtbpy, 4,4′-di-tert-butylbipyridine; r.t., room temperature.

(*34*)—a bicyclic cytotoxic lactone isolated from the fungus *Ajuga decumbens* and recently synthesized by Begum and Chakraborty (*35*)—possesses the bicyclic 5,5-fused scaffold as in lactone **25b** (Fig. 3A). Retrosynthetic disconnection at the lactone C–O bond indicated that complex dicarboxylic acid **61** would be the synthetic intermediate for the key C–H lactonization step. This intermediate (**61**) was prepared efficiently, albeit as a diastereomeric mixture, from commercially available ethyl 2-oxocyclopentane-1-carboxylate (**54**). Specifically, **54** was methylated with MeI in refluxing acetone to provide **55** in quantitative yield, followed by a Wittig reaction with Ph_3PCH_3Br (Ph, phenyl) to produce terminal alkene **56** in 95% yield. The terminal alkene **56** was epoxidized with *meta*-chloroperoxybenzoic acid (*m*CPBA) to generate an inconsequential 1:1 diastereomeric mixture of epoxides (**57**) in 80% yield, which were isomerized with trimethylsilyl trifluoromethanesulfonate (TMSOTf) and 2,6-lutidine in toluene (–78°C to room temperature, overnight) to produce allylic alcohol **58** in 71% yield. The allylic alcohol **58** underwent a Johnson-Claisen rearrangement, followed by olefin isomerization with *p*-toluenesulfonic acid in refluxing toluene to generate the complex cyclopentene **59** in 88% yield. The cyclopentene **59** was hydrogenated with catalytic PtO_2 in acetic acid under the H_2 pressure of a four-layered balloon overnight to provide the complex cyclopentane **60** in quantitative yield as a diastereomeric mixture. Subsequent basic ester hydrolysis of **60** with 15% aqueous NaOH at reflux provided the dicarboxylic acid **61** in quantitative yield. The γ-lactonization of dicarboxylic acid **61** provided the desired lactone **63** in 31% yield, and its diastereomer **62** in 13% yield. Yield calculations for lactones **62** and **63** were based on the reactive diastereomers of **61** (see supplementary materials p. S118 for detailed analysis). The last step of the synthesis was a photocatalytic decarboxylative hydroxylation of **63**, which provided the racemic natural product myrotheciumone A (**64**) in 39% yield (*36*). The total synthesis was completed in nine steps, starting from **54**. Pedicellosine (**65**) (*37*), a phthalide isolated from the leaves of *Gentiana pedicellate* and previously synthesized by Gao and Li (*38*), has the benzo-fused lactone structure, as in lactone **32b** (Fig. 3A). The synthesis began with the preparation of lactone **32b** from the dicarboxylic acid **32** via γ-C–H lactonization, followed by the reduction of the carboxyl group of **32b** to the alcohol using BH_3•Me_2S. The crude alcohol was esterified with 2,3-dihydroxybenzoic acid using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI) as the coupling reagent, providing the simple natural product pedicellosine (**65**) in three steps, starting from **32**. These two representative syntheses demonstrate that the fusion of dicarboxylic acids and C–H lactonization constitutes a viable synthetic strategy amenable to the preparation of molecular targets of varying complexities at both early and late stages of synthesis.

In conclusion, we have achieved site-selective β- and γ-methylene C–H activation of aliphatic acids with two quinoline-pyridone ligands, **L1** and **L2**. This led to the development of three modes of lactonization reaction for the syntheses of γ- and δ-lactones from a variety of dicarboxylic acids. The utility of the methodology is demonstrated by the syntheses of two natural products. Further development of the current methodology for the syntheses of lactones of other ring sizes, as well as extension of the method toward the synthesis of other heterocycles, is underway.

REFERENCES AND NOTES

1. L. C. Kao, A. Sen, *J. Chem. Soc. Chem. Commun.* **1991**, 1242–1243 (1991).
2. B. D. Dangel, J. A. Johnson, D. Sames, *J. Am. Chem. Soc.* **123**, 8149–8150 (2001).
3. R. Giri *et al*., *J. Am. Chem. Soc.* **129**, 3510–3511 (2007).
4. Z. Zhuang *et al*., *J. Am. Chem. Soc.* **140**, 10363–10367 (2018).
5. Z. Zhuang, A. N. Herron, Z. Fan, J. Q. Yu, *J. Am. Chem. Soc.* **142**, 6769–6776 (2020).
6. Z. Zhuang, J. Q. Yu, *Nature* **577**, 656–659 (2020).
7. Z. Zhuang, A. N. Herron, J. Q. Yu, *Angew. Chem. Int. Ed.* **60**, 16382–16387 (2021).
8. K. K. Ghosh, M. van Gemmeren, *Chemistry* **23**, 17697–17700 (2017).
9. K. K. Ghosh, A. Uttry, A. Koldemir, M. Ong, M. van Gemmeren, *Org. Lett.* **21**, 7154–7157 (2019).
10. K. K. Ghosh *et al*., *Angew. Chem. Int. Ed.* **59**, 12848–12852 (2020).
11. F. Ghiringhelli, A. Uttry, K. K. Ghosh, M. van Gemmeren, *Angew. Chem. Int. Ed.* **59**, 23127–23131 (2020).
12. A. Uttry, S. Mal, M. van Gemmeren, *J. Am. Chem. Soc.* **143**, 10895–10901 (2021).
13. Z. Wang *et al*., *Science* **374**, 1281–1285 (2021).
14. M. S. Chen, M. C. White, *Science* **318**, 783–787 (2007).
15. L. Vicens, M. Bietti, M. Costas, *Angew. Chem. Int. Ed.* **60**, 4740–4746 (2021).
16. J. Schwarz, B. König, *Green Chem.* **20**, 323–361 (2018).
17. S. K. Sartori, M. A. N. Diaz, G. Diaz-Muñoz, *Tetrahedron* **84**, 132001–132039 (2021).
18. M. Seitz, O. Reiser, *Curr. Opin. Chem. Biol.* **9**, 285–292 (2005).
19. E. J. Corey, M. C. Kang, M. C. Desai, A. K. Ghosh, I. N. Houpis, *J. Am. Chem. Soc.* **110**, 649–651 (1988).
20. M. T. Crimmins *et al*., *J. Am. Chem. Soc.* **122**, 8453–8463 (2000).
21. E. J. Corey, W. G. Su, *J. Am. Chem. Soc.* **109**, 7534–7536 (1987).
22. M. A. Baker, R. M. Demoret, M. Ohtawa, R. A. Shenvi, *Nature* **575**, 643–646 (2019).
23. E. J. Corey, H. L. Pearce, *J. Am. Chem. Soc.* **101**, 5841–5843 (1979).
24. B. M. Trost, M. J. Krische, *J. Am. Chem. Soc.* **118**, 233–234 (1996).
25. S. W. M. Crossley, G. Tong, M. J. Lambrecht, H. E. Burdge, R. A. Shenvi, *J. Am. Chem. Soc.* **142**, 11376–11381 (2020).
26. E. Lee *et al*., *J. Am. Chem. Soc.* **119**, 8391–8392 (1997).
27. G. Shabir, A. Saeed, H. R. El-Seedi, *Phytochemistry* **181**, 112568–112787 (2021).
28. A. D. Kinghorn, H. Falk, S. Gibbons, J. Kobayashi, Eds., *Progress in the Chemistry of Organic Natural Products* (Springer Nature, 2017), vol. 104.
29. F. Zaragoza Dörwald, *Side Reactions in Organic Synthesis: A Guide to Successful Synthesis Design* (Wiley, 2004).
30. X. Chen, J. J. Li, X. S. Hao, C. E. Goodhue, J. Q. Yu, *J. Am. Chem. Soc.* **128**, 78–79 (2006).
31. G. Xia *et al*., *Nat. Chem.* **11**, 571–577 (2019).
32. P. Novák, A. Correa, J. Gallardo-Donaire, R. Martin, *Angew. Chem. Int. Ed.* **50**, 12236–12239 (2011).
33. J. M. Lee, S. Chang, *Tetrahedron Lett.* **47**, 1375–1379 (2006).
34. T. Lin *et al*., *Bioorg. Med. Chem. Lett.* **24**, 2504–2507 (2014).
35. S. Begum, T. K. Chakraborty, *J. Org. Chem.* **86**, 11812–11821 (2021).
36. S. N. Khan, M. K. Zaman, R. Li, Z. Sun, *J. Org. Chem.* **85**, 5019–5026 (2020).
37. E. M. Mpondo, J. Garcia, A. J. Chulia, *Planta Med.* **53**, 297–298 (1987).
38. Q. Gao, S. B. Li, *Chin. Chem. Lett.* **10**, 109–110 (1999).

ACKNOWLEDGMENTS

We thank D. Strassfeld, M. Tomanik, and K. Wu for proofreading and providing helpful suggestions in preparing the manuscript; N. Lam for helpful discussions; and Z. Wang for the preparation of ligands. **Funding:** We acknowledge The Scripps Research Institute, NIH (NIGMS, 2R01GM084019), and Croucher Foundation (postdoctoral fellowship to H.S.S.C.) for financial support. **Author contributions:** J.-Q.Y. conceived the study concept. H.S.S.C. discovered and developed the three modes of lactonization reactions and the use of silver-free conditions, conducted preliminary mechanistic studies, and finished the total synthesis of myrotheciumone A. J.-M.Y. developed the substrate scope for the benzo-fused lactones and part of the substrate scope for the aliphatic lactones and finished the synthesis of pedicellosine. H.S.S.C. and J.-Q.Y. wrote the manuscript. J.-Q.Y. directed the project. **Competing interests:** J.-Q.Y. and H.S.S.C. are inventors on a patent application related to this work (US Patent application 63/339,627) filed by The Scripps Research Institute. The authors declare no other competing interests. **Data and materials availability:** All data are in the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq3048
Materials and Methods
Figs. S1 to S10
Tables S1 to S15
NMR Spectra
References (*39*–*45*)

Submitted 30 March 2022; accepted 17 May 2022
Published online 26 May 2022
10.1126/science.abq3048

MARINE LIPIDS

Global ocean lipidomes show a universal relationship between temperature and lipid unsaturation

Henry C. Holm[1,2], Helen F. Fredricks[1], Shavonna M. Bent[1,2], Daniel P. Lowenstein[1,2], Justin E. Ossolinski[1], Kevin W. Becker[1]†, Winifred M. Johnson[1,2]‡, Kharis Schrage[1,2], Benjamin A. S. Van Mooy[1]*

Global-scale surveys of plankton communities using "omics" techniques have revolutionized our understanding of the ocean. Lipidomics has demonstrated the potential to add further essential insights on ocean ecosystem function but has yet to be applied on a global scale. We analyzed 930 lipid samples across the global ocean using a uniform high-resolution accurate-mass mass spectrometry analytical workflow, revealing previously unknown characteristics of ocean planktonic lipidomes. Focusing on 10 molecularly diverse glycerolipid classes, we identified 1151 distinct lipid species, finding that fatty acid unsaturation (i.e., number of carbon-carbon double bonds) is fundamentally constrained by temperature. We predict substantial declines in the essential fatty acid eicosapentaenoic acid over the next century, which are likely to have serious deleterious effects on economically critical fisheries.

Lipids are a class of biomolecules produced and used by organisms from all domains of life for energy storage, membrane structure, and signaling. Lipids make up 10 to 20% of the particulate organic carbon in the surface ocean where lipid production and inventories are greatest (*1*–*3*). For decades, oceanographers have used lipids as biomarkers of chemical and biological ocean processes (*4*). Despite robust research into their biogeochemistry, the combination of high-resolution mass spectrometry and downstream analytical tools has only recently allowed for comprehensive untargeted assessments of ocean lipids, on scales akin to surveys of other molecules such as nucleic acids and proteins (*5*–*8*). These new tools allow for the examination of hundreds to thousands of lipid species in a sample (the entirety of which is referred to as the "lipidome") and present the opportunity to holistically examine global factors affecting ocean lipid composition. Critically, marine plankton lipidomes are likely to change as a function of water temperature, dissolved nutrient concentration, salinity, community composition, and other properties of the surface ocean. Lipid membranes are a core part of cellular adaptions to environmental perturbations, making lipidomics an important tool for understanding how planktonic communities in the surface ocean will shift as a result of climate change (*9*).

We present here a global-scale mass spectral dataset of planktonic lipidomes from 146 locations (Fig. 1A) with concurrent environmental metadata aggregated from seven oceanographic research cruises using standardized collection, preservation, and extraction methods. We used high-performance liquid chromatography coupled with electrospray ionization high-resolution accurate-mass mass spectrometry (HPLC-ESI-HRAM-MS) to generate spectra from 930 lipid extracts from samples of planktonic particles retained on 0.22-μm pore-size membranes. Using the R coding language packages XCMS, CAMERA, and LOBSTAHS, we searched this spectral dataset for the 10 dominant glycerolipid molecular classes found in surface ocean particulate organic carbon (*3*, *5*, *10*); our methods are not optimized for archaeal ether lipids and, thus, they are not included in our study. We retained 1151 high-confidence intact glycerolipid annotations based on a combination of accurate mass, MS/MS spectra, adduct hierarchy, retention time, and positive- and negative-mode ionization characteristics (*11*). Lipids from each class were quantified by using external curves of a representative lipid species. This approach has been validated by our laboratory for possible matrix effects through isotopically labeled internal standards (*3*).

We defined a lipid species' relative abundance as the mass percent of the total lipidome. Within the dataset, annotated glycerolipids displayed a unimodal distribution of abundances in logarithmic space with a small number of species making up the majority of the annotated lipid mass (Fig. 1B). We defined a lipid species' prevalence in the dataset as the percentage of samples in which the species was observed. Aside from an abundant subset of species present in >90% of the samples, prevalence across the dataset was fairly evenly distributed with a similar numbers of lipid

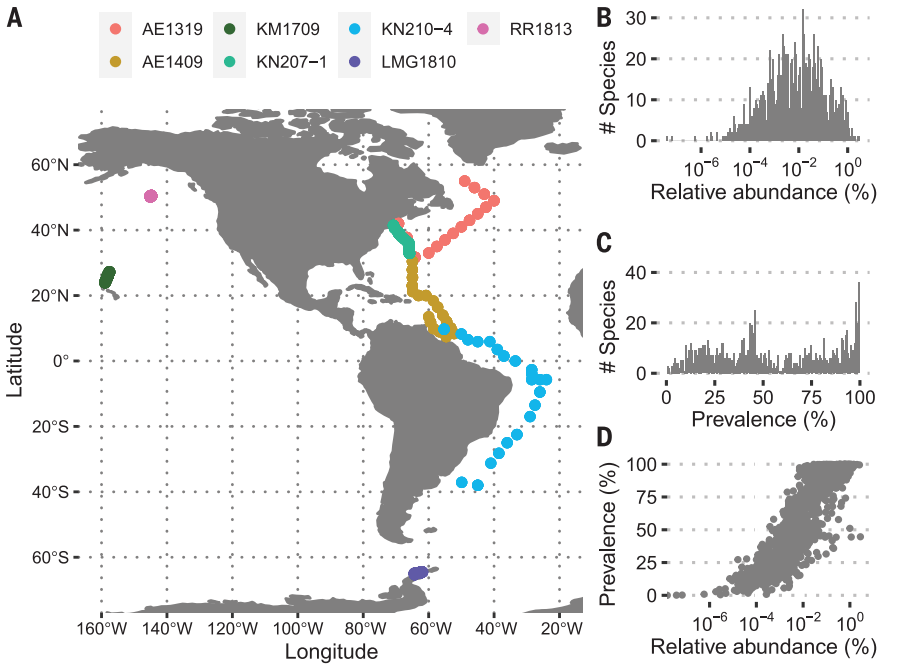


Fig. 1. Compositional structure of glycerolipid species in the global ocean lipidome. (**A**) A map of dataset sampling locations with sites colored by cruise. (**B**) Histogram of lipid species relative abundance in dataset. Relative abundance for each lipid species is the percentage of the total mass of all annotated lipids. (**C**) Histogram of lipid species prevalence in dataset. Prevalence is defined as the percentage of samples in which a lipid species is found. (**D**) The relationship between prevalence and relative abundance for each lipid species.

[1]Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution (WHOI), Woods Hole, MA 02543, USA. [2]Massachusetts Institute of Technology (MIT)–WHOI Joint Program in Oceanography/Applied Ocean Science and Engineering, Cambridge, MA 02139, USA.
*Corresponding author. Email: bvanmooy@whoi.edu
†Present Address: GEOMAR Helmholtz Centre for Ocean Research Kiel, 24105 Kiel, Germany. ‡Present Address: Department of Chemistry and Biochemistry, University of North Carolina Wilmington, Wilmington, NC 28403, USA.

species across lower percentages (Fig. 1C). Both these metrics were positively correlated, with higher-abundance lipids being more prevalent (Fig. 1D). Owing to this relationship, the most abundant 247 of the 1151 lipid species represented 90% of the total lipid mass. Of these highly abundant species, the majority are present in more than 97% of samples. These metrics show that although molecular diversity in the ocean is relatively high, a comparably small number of highly common lipid species make up an outsized proportion of the mass. In total, this dataset represents more than 600,000 identified peaks, but numerous other mass spectral features remained to be conclusively annotated. The raw spectral files are publicly available to serve as a resource for continued global ocean investigations similar to that of other large "omics"-based field studies.

Planktonic community lipidomes are affected by numerous environmental factors, e.g., nutrient availability (*12*), but here we report on the relationship between lipids and arguably the most fundamental control on their composition: temperature (*13*). In a process called homeoviscous adaptation, organisms adapt to temperature-induced changes in membrane fluidity through shifts in the unsaturation level (i.e., number of double bonds) of their fatty acid moieties (*13*, *14*). Our approach allowed us to examine the unsaturation state of individual glycerolipid classes in natural marine microbial communities across the oceans and differentiate class-internal temperature-induced shifts in unsaturation from shifts between different classes of distinct unsaturation (*15*). Throughout the ocean, the surface mixed layer typically has the highest lipid concentrations and the steepest latitudinal temperature gradients (*1*). Therefore, of the total 930 samples spanning 0- to 600-m depth, we analyzed the 243 samples collected from the surface mixed layer (generally <40 m; materials and methods 1.1).

In Fig. 2, we present the saturation state in the mixed layer for four representative glycerolipid classes: (i) sulfoquinovosyl diacylglycerols (SQDG, Fig. 2A), glycolipids found almost exclusively in eukaryotic photosynthetic membranes and cyanobacteria (*10*); (ii) triacylglycerols (TAG, Fig. 2B), energy-storage lipids from eukaryotic plankton, particularly those under nutrient stress, and the most abundant glycerolipid that we identified (*3*); (iii) diacylglyceryl trimethylhomoserines and diacylglyceryl hydroxymethyltrimethyl-β-alanines (DGTS/DGTA, Fig. 2C), betaine lipids important for adaption to nutrient stress with broad distribution among marine plankton (*16*); and (iv) phosphatidylethanolamines (PE, Fig. 2D), one of the major phospholipids found in membranes of heterotrophic bacteria (*10*). Among these glycerolipid classes, we found temperature to be highly influential in structuring the

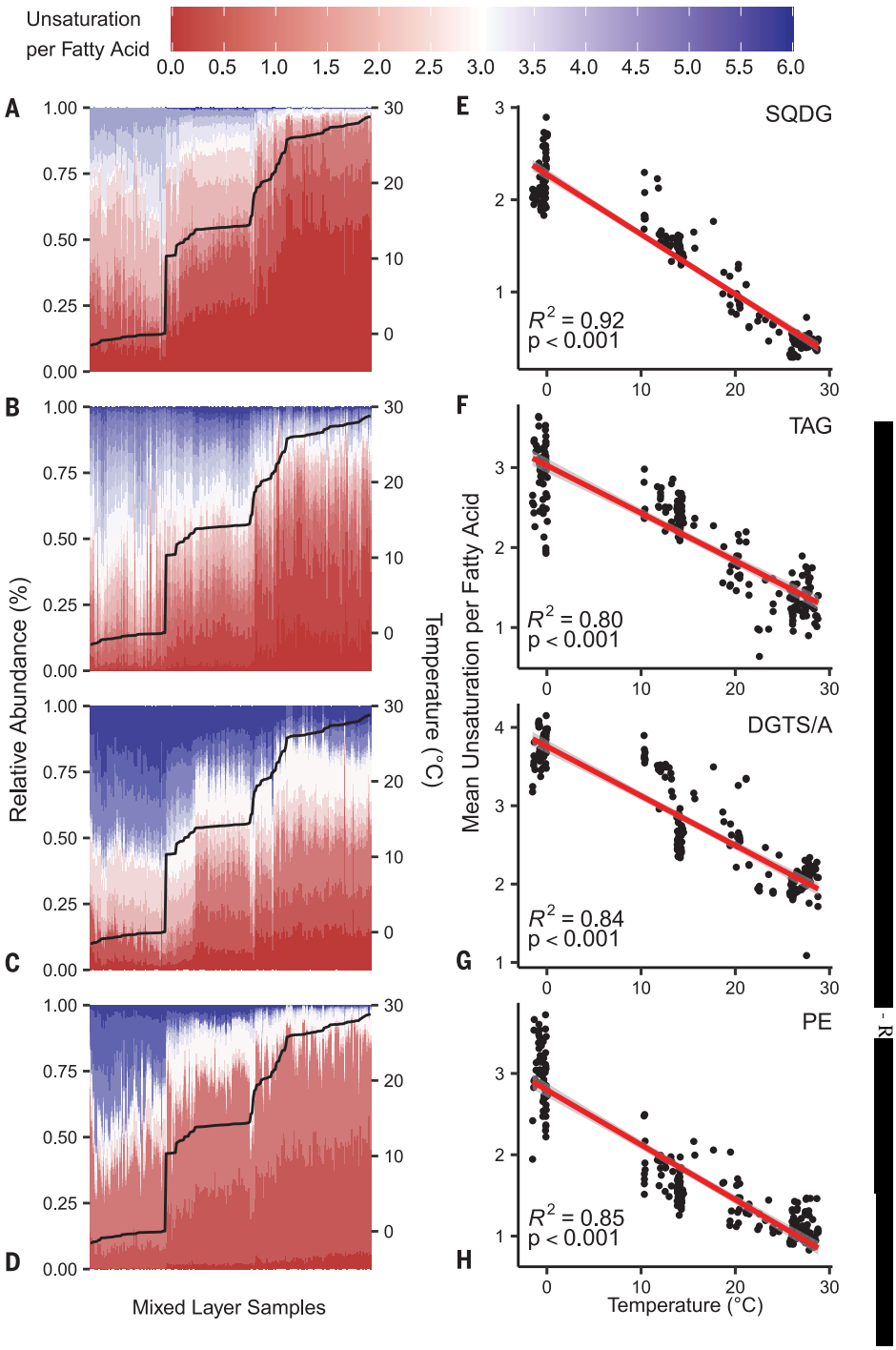


Fig. 2. Mixed-layer temperature is highly correlated with suspended fatty acid unsaturation. Four representative glycerolipid classes are shown: a glycolipid (SQDG) (A and E), an energy storage lipid (TAG) (B and F), betaine lipids (DGTS/DGTA) (C and G), and a phospholipid (PE) (D and H). Identical figures for the other glycerolipids are provided in figs. S1 and S2. (**A** to **D**) Stacked bar plots show the relative abundance within each class of lipid species in mixed layer samples. Water temperature for each sample is shown with a black line plotted over the stacked bar plot (right axis). Samples are sorted from coldest to warmest along the *x* axis. Lipid species are sorted from least (red) to most (blue) unsaturated along the *y* axis and are colored to indicate the average number of unsaturations per fatty acid moiety on the glycerolipid. (**E** to **H**) Weighted mean number of unsaturations for each lipid class versus temperature among mixed-layer samples. Red line shows linear fits; inset shows coefficient of determination for linear regression with ordinary least squares and *p* value from F-test (n = 242 samples). Gray range shows 95% confidence interval of fit based on SE.

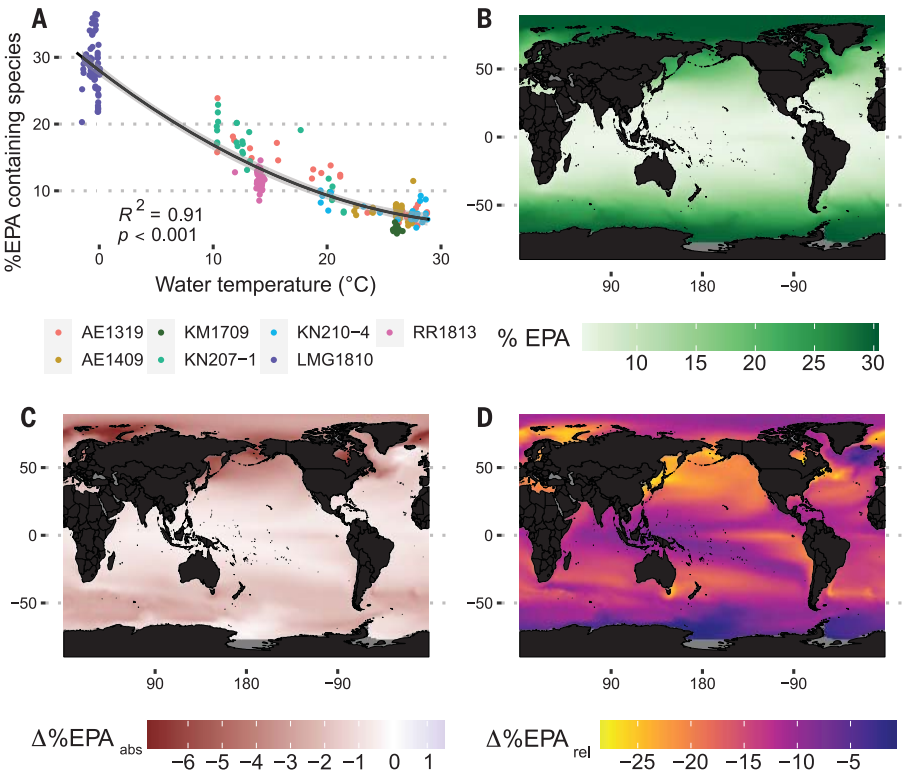


Fig. 3. Current and future projections of EPA abundance based on water temperature shows loss under high-emission (SSP5-8.5) scenario. (**A**) Percent abundance of all polar glycerolipids containing EPA species (% EPA) in mixed layer samples verse temperature. Samples colored by cruise. Black line shows quadratic fit, gray area shows 99% confidence interval using SE. Statistics for the fit with least squares are shown in plot (n = 242 samples, p values from F-test). (**B**) Current %EPA in mixed layer predicted using CMIP6 multimodel mean historical SST (1995 to 2014). (**C**) Change in absolute %EPA ($\Delta\%EPA_{abs}$) at end of century from today using SST under SSP5-8.5 scenario in the same models (2080 to 2099). (**D**) Projected relative loss of EPA ($\Delta\%EPA_{rel}$).

relative abundance of fatty acid species; there is a clear transition from species with more unsaturated fatty acids at colder temperatures to fully saturated species (i.e., with only single bonds) at the warmest temperatures (Fig. 2, A to D). These trends are also evident in all the other glycerolipid classes (fig. S1), as well as in the total aggregated lipidome of all glycerolipid classes (fig. S2).

From this compositional information, we calculated a weighted average unsaturation state for each glycerolipid class. For all the glycerolipid classes, this weighted mean unsaturation derived from the composition was highly linearly correlated to water temperature (Fig. 2, E to H, and table S1). All classes showed shifts along this gradient, a finding that is notable given both the different inherent unsaturation states, taxonomic sources, and biochemical roles of each lipid class (e.g., SQDG in thylakoid membranes, DGTS/DGTA and PE in cellular membranes, and TAGs in lipid bodies). There is considerable scatter in the data at low temperatures collected off the Antarctic coast, which might be due to mixing between truly planktonic organisms and plankton released from melting sea ice. Using a linear fit to the weighted mean unsaturation of all lipid classes combined, we found that the number of unsaturations nearly triples across the temperature range sampled (1.2 unsaturations per fatty acid at 29°C, versus 3.3 at −2°C; fig. S2 and table S1).

Given the tight correlation of unsaturation to temperature within individual lipid classes across highly diverse geographic areas, we suggest that this trend stems from a universal biophysical necessity to control cell membrane fluidity through homeoviscous adaptation. Although our data are not sufficient to absolutely implicate this mechanism, the consistency of temperature as a predictor necessitates a fundamental biophysical explanation. Even when accounting for other environmental parameters (depth, salinity, nutrients, light, and time of year) in a linear multiple regression analysis, temperature is the most powerful explanatory variable for membrane lipid unsaturation (fig. S3 and materials and methods 1.8). The genetic ability for homeoviscous adaption appears to be widely distributed; in cell culture, a wide range of marine phytoplankton taxa have been shown to change their membrane unsaturation with changing temperature (*17*–*19*). Our study transects numerous ocean biomes and cannot resolve the question of whether temperature itself or links between temperature and planktonic community composition drive the trends that we observe (*20*), but future "omics" studies could shed light on this; e.g., a metatranscriptomic survey of fatty acid desaturases. Indeed, it is notable that the relationship between temperature and unsaturation emerges from our dataset despite spanning such diverse and disparate planktonic communities, from the nutrient-depleted subtropical gyres to the highly productive Antarctic coastal shelf.

Within the compositional changes seen in this dataset, we were interested in how the abundance of two specific unsaturated fatty acids—eicosapentaenoic acid (EPA 20:5n-3) and docosahexaenoic acid (DHA 22:6n-3)—shifted along the temperature gradient. These fatty acids are essential for nutrition in zooplankton and other higher–trophic level organisms and must be obtained largely from their diet (*21*). Thus, phytoplankton are the primary source of these lipids in marine food webs (*22*). They are often referred to as long-chain essential fatty acids (LCEFAs; along with α-linoleic acid 18:3n-3). Using diagnostic MS/MS spectra, we were able to determine whether a mass feature contained a LCEFA based on characteristic fatty acid fragments in the intact polar glycerolipids, and we singled these out for further exploration (11). These fragments were difficult to deconvolute in TAGs, limiting definitive LCEFA detection, although species containing more than five double bonds are readily identifiable (fig. S4). We found that the percent abundance of all polar glycerolipids (i.e., excluding TAGs) containing a LCEFA in the mixed layer varied between 3 and 36% and 6 and 28% for EPA and DHA, respectively (Fig. 3A and fig. S5). Whereas the percent abundance of EPA species showed a strong relationship to temperature, the correlations for lipids with DHA was weaker. In addition to affecting membrane fluidity, DHA has also been shown to affect permeability and membrane fusion more than other unsaturated species do (*23*). Thus, our results may indicate that planktonic organisms use EPA for regulating the phase transition temperature of membranes and DHA for other functions.

Given the strong relationship between planktonic EPA abundance and temperature, we sought to examine current geospatial patterns of this LCEFA. We fit a quadratic curve (R^2 = 0.91, p < 0.01, n = 239; table S2) to the percent abundance of polar glycerolipids containing EPA species (%EPA) versus temperature in mixed-layer samples (Fig. 3A) and then used

sea-surface temperature (SST) data from the model intercomparison project "CMIP6" to predict EPA abundance in the mixed layer [Fig. 3, B to D; modeled SST data from Kwiatkowski *et al.* 2020 (*24*)]. These temperature values come from 13 combined Earth system models for historical and future temperature scenarios. Using the multimodel mean annual average SST temperatures, we project that high-latitude waters contain proportionally much more EPA than low-latitude areas (Fig. 3B). On the basis solely of temperature, we found the mean % EPA between 50° and 60° latitude (19.4 ± 3.41% for 50° to 60°N, 22.9 ± 2.61% for 50° to 60°S) to be four times the amount between 10° and 0° latitude (5.8 ± 0.32% for 10° to 0°N,6.0 ± 0.46% for 10° to 0°S). The seasonal range in these %EPA estimates could be even greater. To see how the upper and lower limits for this composition are likely to shift under future warming conditions, we generated %EPA maps using end-of-century SST conditions from SSP1-2.6 and SSP5-8.5 scenarios in the same multimodel (Fig. 3, C and D, and fig. S6). SSP1-2.6 represents a high-mitigation scenario with low CO_2 emission, whereas SSP5-8.5 represents a higher-CO_2 emission scenario. Under the SSP5-8.5 scenario, which, admittedly, is a worst-case scenario (*25*), the change in mean absolute %EPA ($\Delta\%EPA_{abs} = \%EPA_{future} - \%EPA_{present}$) was about −2% globally, but with some areas, particularly at higher latitudes, seeing a more drastic decrease of up to −7% $\Delta\%EPA_{abs}$ (Fig. 3C). The largest relative changes ($\Delta\%EPA_{rel} = (\Delta\%EPA_{abs}/\%EPA_{present}) \times 100$) are more geographically heterogeneous, ranging from -0.7% to -28.5%. For example, locations such as the Sea of Japan, Norway Sea, and Grand Banks lose up to a quarter of current EPA amounts (Fig. 3D). In addition, broad expanses of the western North Atlantic and western North Pacific also showed high relative decreases.

These observations of EPA's correlation to water temperature and future projections of its decline in abundance suggest global threats to this LCEFA's availability to higher-trophic levels by the end of this century. Previous research in zooplankton (e.g., copepods) has shown that some species experience large swings in their LCEFA content when fed different diets that vary in prey fatty acid composition (*26*). Furthermore, changes in the abundance of other precursor unsaturated fatty acids may further exacerbate stress from dietary LCEFA deficiency. Specifically, LCEFA-starved invertebrates may express desaturases and elongases to synthesize LCEFAs from 18:3n-3, 18:3n-6, and 18:4n-3 fatty acids; this process is very inefficient compared to dietary EPA acquisition, and future decreases in the availability of these precursor fatty acids would place an additional biochemical tax on EPA production (*27*). In many fish species, diets that are poor in EPA lead to deleterious changes in immune function and therefore fitness (*28*). Thus, it is possible that declines in primary producer EPA levels could affect both nutritional value of fish (e.g., for human consumption) and their overall population stability. Specifically, fisheries in the aforementioned Sea of Japan, Norwegian Sea, and Grand Banks, as well as coastal fisheries around Alaska, eastern Russia, and the Peruvian upwelling system, all face high projected losses in relative amounts of EPA, adding another layer of stress to already projected shifts in fishery baselines (*29*). Additionally, the influences of projected warming and retreating ice cover on LCEFA shifts in the Arctic Ocean are unknown but likely to be even more substantial.

Better understanding of environmental factors and mechanisms affecting the planktonic community lipidomes is critical to projecting future changes in ocean ecosystem services. Previous studies have shown that rising temperatures are likely to cause poleward shifts in planktonic thermal niches (*30*). The expansion of gyres and permanent changes to phytoplankton communities may alter the fatty acid composition of an ocean province year-round. However, with fatty acid unsaturation tied to rapid adaptation mechanisms, it is possible that marine heat waves could immediately affect phytoplankton lipids on much shorter time scales. This raises the possibility of transient events of decreased unsaturated lipid production with unknown effects on the life histories of economically important taxa or their prey. Given that the frequency and intensity of short-lived marine heat waves are projected to increase, more research is needed into how this may affect fatty acid unsaturation on daily to seasonal time scales (*31*).

REFERENCES AND NOTES

1. S. G. Wakeham, C. Lee, J. I. Hedges, P. J. Hernes, M. J. Peterson, *Geochim. Cosmochim. Acta* **61**, 5363–5369 (1997).
2. B. Gašparović, A. Penezić, R. S. Lampitt, N. Sudasinghe, T. Schaub, *Mar. Chem.* **186**, 101–109 (2016).
3. K. W. Becker *et al.*, *Nat. Commun.* **9**, 5179 (2018).
4. J. P. Koelmel *et al.*, *Metabolomics* **16**, 56 (2020).
5. J. R. Collins, B. R. Edwards, H. F. Fredricks, B. A. S. Van Mooy, *Anal. Chem.* **88**, 7154–7162 (2016).
6. M. A. Saito *et al.*, *Science* **345**, 1173–1177 (2014).
7. S. Sunagawa *et al.*, *Nat. Rev. Microbiol.* **18**, 428–445 (2020).
8. S. I. Cantarero *et al.*, *Front. Mar. Sci.* **7**, 540643 (2020).
9. I. Marinov, S. C. Doney, I. D. Lima, *Biogeosciences* **7**, 3941–3959 (2010).
10. K. J. Popendorf, M. W. Lomas, B. A. S. Van Mooy, *Org. Geochem.* **42**, 803–811 (2011).
11. Materials and methods are available as supplementary materials.
12. B. A. S. Van Mooy *et al.*, *Nature* **458**, 69–72 (2009).
13. M. Sinensky, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 522–525 (1974).
14. J. R. Hazel, E. E. Williams, *Prog. Lipid Res.* **29**, 167–227 (1990).
15. B. A. S. Van Mooy, H. F. Fredricks, *Geochim. Cosmochim. Acta* **74**, 6499–6516 (2010).
16. J. P. Cañavate, I. Armada, J. L. Ríos, I. Hachero-Cruzado, *Phytochemistry* **124**, 68–78 (2016).
17. H. Jiang, K. Gao, *J. Phycol.* **40**, 651–654 (2004).
18. P. A. Thompson, M. Guo, P. J. Harrison, J. N. C. Whyte, *J. Phycol.* **28**, 488–497 (1992).
19. J. Pittera *et al.*, *Environ. Microbiol.* **20**, 612–631 (2018).
20. A. W. E. Galloway, M. Winder, *PLOS ONE* **10**, e0130053 (2015).
21. L. F. C. Castro, D. R. Tocher, O. Monroig, *Prog. Lipid Res.* **62**, 25–40 (2016).
22. L. Ruess, D. C. Müller-Navarra, *Front. Ecol. Evol.* **7**, 269 (2019).
23. W. Ehringer, D. Belcher, S. R. Wassall, W. Stillwell, *Chem. Phys. Lipids* **54**, 79–88 (1990).
24. L. Kwiatkowski *et al.*, *Biogeosciences* **17**, 3439–3470 (2020).
25. Z. Hausfather, G. P. Peters, *Nature* **577**, 618–620 (2020).
26. A. J. Veloza, F.-L. E. Chu, K. W. Tang, *Mar. Biol.* **148**, 779–788 (2006).
27. Ó. Monroig, N. Kabeya, *Fish. Sci.* **84**, 911–928 (2018).
28. M. T. Arts, M. T. Brett, M. J. Kainz, Eds., "Chapter 10, Health and Condition in Fish: The Influence of Lipids on Membrane Competency and Immune Response" in *Lipids in aquatic ecosystems* (Springer, 2009). pp. 237–255.
29. W. W. L. Cheung *et al.*, *Glob. Change Biol.* **16**, 24–35 (2010).
30. M. K. Thomas, C. T. Kremer, C. A. Klausmeier, E. Litchman, *Science* **338**, 1085–1088 (2012).
31. T. L. Frölicher, E. M. Fischer, N. Gruber, *Nature* **560**, 360–364 (2018).
32. H. Holm, hholm/OceanLipidome: Version 1.0.1, Zenodo (2022). https://doi.org/10.5281/zenodo.6394224.
33. M. W. Lomas, A. Martiny, Depth profile data from R/V Atlantic Explorer AE1319 in the NW Atlantic from Aug-Sept. 2013", 2020-11-2, Woods Hole Open Access Server (WHOAS) (2020). https://doi.org/10.26008/1912/bco-dmo.829797.1.
34. E. Kujawinski, K. Longnecker, Processed CTD data from all sensors mounted on the rosette from R/V Knorr cruise KN210-04 in the Western Atlantic Ocean between Uruguay and Barbados in 2013 (Deep Atlantic DOM project), Biological and Chemical Oceanography Data Management Office (BCO-DMO) (2014). https://doi.org/10.1575/1912/bco-dmo.481164.1.
35. B. A. Van Mooy, CTD profiles from R/V Knorr cruise KN207-01 in the southern tip of Nova Scotia to Bermuda in 2012 (SargassoSeaLipids project). Biological and Chemical Oceanography Data Management Office (BCO-DMO) (2012). https://lod.bco-dmo.org/id/dataset/3666.
36. B. A. Van Mooy, S. T. Dyhrman, (2015) CTD data from R/V Atlantic Explorer cruise AE1409 in the Western Tropical North Atlantic in 2014 (P Processing by Tricho project). Biological and Chemical Oceanography Data Management Office (BCO-DMO) (2015). https://lod.bco-dmo.org/id/dataset/616342.
37. B. Van Mooy, CTD (Sea-Bird SBE-911+) data as collected during the cruise LMG1810, Rolling Deck to Repository (R^2R) (2021). https://doi.org/10.7284/138731.
38. RR1813 cruise data, https://doi.org/10.5067/SeaBASS/EXPORTS/DATA001.
39. KM1709 data, http://scope.soest.hawaii.edu/data/mesoscope/mesoscope.html.

ACKNOWLEDGMENTS

We thank the captains and the crews of the RVs *Knorr, Atlantic Explorer, Kilo Moana, Roger Revelle,* and *Lawrence M. Gould* as well as chief scientists M. Lomas (AE1319), B. Barone and T. Clemente (KM1709), E. Kujawinski (KN210-4), and D. Steinberg (RR1813). L. Kwiatkowski graciously provided multimodel means of SST used here from multiple CMIP6 model projections. K. Perry and M. Small processed and prepared lipid samples from LMG1810. J. Tagliaferre collected samples on AE1319. A. Santoro and P. Morton collected lipid samples for the RR1813 cruise at sea. F. Hopewell collected lipid samples on cruise KM1709. **Funding:** This research was supported by grants to B.A.S.V.M, from National Science Foundation grant OCE-1031143, National Science Foundation grant OCE-1045670, National Science Foundation grant OCE-1059884, National Science Foundation grant OCE-1332898, National Science Foundation grant OPP-1543328, National Science Foundation grant OCE-1756254, National Science Foundation grant OCE-2022597, National Science Foundation grant OPP-2026045, The Marine Microbiology Initiative division of the Gordon and Betty Moore Foundation grant 5073, and Simons Foundation grant 721229. **Author contributions:** Project administration: B.A.S.V.M. Conceptualization: B.A.S.V.M. Funding acquisition: B.A.S.V.M. Data curation: H.C.H. Formal analysis: H.C.H. Investigation: H.C.H., H.F.F.,

S.M.B., D.P.L., J.E.O., K.W.B., W.M.J., J.P.T., K.R.S. Methodology: B.A.S.V.M., H.C.H., H.F.F. Visualization: H.C.H. Writing – original draft: H.C.H., B.A.S.V.M. Writing – review & editing: H.C.H., H.F.F., S.M.B., D.P.L., J.E.O., K.W.B., W.M.J., K.R.S., B.A.S.V.M. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data from this analysis including lipid annotations and applicable code is free to access on Zenodo [(*32*), version 1.0.1]. The unprocessed mass spectra for each lipid sample in this study are available via the *Metabolights* repository alongside relevant environmental metadata for each sample (https://www.ebi.ac.uk/metabolights/MTBLS2838). Modeled SST values from Kwiatkowski *et al.* are available online from the CMIP6 repository [(*24*), DOE node: https://esgf-node.llnl.gov/search/cmip6/]. All CTD data are additionally available from BCO-DMO (https://www.bco-dmo.org/) for four cruises (*33*–*36*), Rolling Deck to Repository (*37*) for LMG1810, on SeaBass for RR1813 (*38*), and at the SCOPE collaboration website (*39*) for KM1709. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn7455
Materials and Methods
Figs. S1 to S11
Tables S1 to S6
References (*40*–*53*)

View/request a protocol for this paper from *Bio-protocol*.

Submitted 17 December 2021; accepted 29 April 2022
10.1126/science.abn7455

FOREST ECOLOGY

8000-year doubling of Midwestern forest biomass driven by population- and biome-scale processes

A. M. Raiho[1,2]*, C. J. Paciorek[3], A. Dawson[4,5], S. T. Jackson[6,7], D. J. Mladenoff[8], J. W. Williams[9,10], J. S. McLachlan[1]*

Changes in woody biomass over centuries to millennia are poorly known, leaving unclear the magnitude of terrestrial carbon fluxes before industrial-era disturbance. Here, we statistically reconstructed changes in woody biomass across the upper Midwestern region of the United States over the past 10,000 years using a Bayesian model calibrated to preindustrial forest biomass estimates and fossil pollen records. After an initial postglacial decline, woody biomass nearly doubled during the past 8000 years, sequestering 1800 teragrams. This steady accumulation of carbon was driven by two separate ecological responses to regionally changing climate: the spread of forested biomes and the population expansion of high-biomass tree species within forests. What took millennia to accumulate took less than two centuries to remove: Industrial-era logging and agriculture have erased this carbon accumulation.

Woody biomass contains one of the largest pools of terrestrial carbon. Its long-term trajectory is paced by changes in forest composition and structure over centuries to millennia (*1*), but most direct observations of vegetation biomass span no more than a few decades. Therefore, estimates of the long-term dynamics of woody biomass, such as those in simulation models (*2*), remain underconstrained by data. High uncertainty about the strength and pacing of the long-term terrestrial carbon sink leads to uncertainties in projections of the future carbon climate system (*3*, *4*).

In the absence of empirical estimates of long-term trends in vegetation biomass, researchers have routinely assumed that ecosystems were in some form of steady state at broad scales before the disruptions imposed by industrial-era land transformation (*5*). This expectation provides a benchmark against which industrial-era changes in vegetation are evaluated (*6*), and it justifies the common practice of spinning up preindustrial vegetation to equilibrium conditions in simulation models [e.g., (*7, 8*)]. This assumption is known to be imperfect, given variations in forest composition over the preindustrial Holocene [PIH, 11,700 years ago to 1850 CE] documented by networks of fossil pollen records (*9*). However, the extent to which these documented changes in vegetation composition imply changes in woody carbon pools is poorly understood because vegetation biomass has a complex relationship with the composition of fossil pollen assemblages.

Understanding the details of how long-term changes in vegetation biomass emerge from changes in vegetation composition is important because competing representations of this relationship have substantial implications for our understanding of the pattern and pace of terrestrial carbon storage and release. For example, the most common current approach to reconstructing woody biomass from fossil pollen data networks is to reconstruct changes in the spatial distribution of biomes (*10*), to which estimates of carbon density are assigned (*11*, *12*). Biomes are treated monolithically in these reconstructions (*10*), so changes in biomass only occur as biome distributions shift over time. The simplifying assumption that shifts in vegetation at the biome scale sufficiently characterize long-term trajectories of woody biomass is challenged by observations that sub-biome variation in plant populations and communities has a large influence on woody-biomass pools (*13*) and that these variations can be uncorrelated with biome-scale change over time (*9*). Simulations suggest that changes in carbon pools within biomes can approach the magnitude of those associated with biome shifts (*8*, *14*).

Previous research has leveraged the population- and community-level information in fossil pollen data to reconstruct localized biomass variations over millennia, but these studies have been challenged by difficulties accounting for uncertainties in the complex relationship between pollen and vegetation. For example, local-scale reconstructions of Holocene variations in tree biomass based on the accumulation rates of pollen or plant macrofossils (*15*—*17*) are sensitive to within- and among-basin variation in sediment influx and do not readily scale up to broader spatial scales. Ongoing efforts continue to improve our characterization of statistical and process uncertainties in pollen-based biomass reconstructions (*18*). The one previous study to attempt broad-scale reconstructions of forest biomass was based on transfer functions calibrated to satellite-derived biomass estimates and suggested no change in total aboveground biomass in eastern Canada over the past 4000 years, apparently upholding the common assumption of a steady-state carbon cycle before the onset of the industrial revolution (*19*).

Here, we provide a statistically robust reconstruction, at centennial resolution, of aboveground woody biomass (AGWB, in megagrams of dry biomass per hectare) across the PIH for a >600,000-km^2 study area in the upper Midwestern region of the United States, a region with well-documented Holocene changes in vegetation composition and distribution at ecological levels from species to biomes. ReFAB (Reconstructing Forest Aboveground Biomass) is a Bayesian statistical model that makes temporally smoothed estimates of AGWB at 100-year intervals from time series of fossil pollen assemblages in sediments. ReFAB accounts for temporal autocorrelation, uncertainty in sediment dating, and uncertainty in the relationship between AGWB and multivariate pollen data (*20*). We calibrated the relationship between AGWB during the 19th century across the region [Fig. 1A (*21*)] and contemporaneous fossil pollen samples. We used historical forest survey data from the 19th century to calibrate ReFAB because these data were collected before the massive changes in vegetation structure and composition caused by regional deforestation and land clearance in the late 19th and early 20th centuries (*22*). The 19th-century vegetation therefore provides better analogs for Holocene vegetation than do contemporary forests (*20*, *23*).

As expected given prior work (*23*), the relationship between the pollen composition of sediments and vegetation across the landscape is complex and nonlinear (e.g., fig. S7). Although site-level predictions of AGWB at the 154 sites used to fit our calibration model show high uncertainty (blue circles in Fig. 1B; R^2 = 0.447), the predicted AGWB across sites fell along the 1:1 observation-to-prediction line (blue line in Fig. 1B), indicating that although there is uncertainty in our model's ability to predict AGWB at any individual site, the model's predictions across sites are not biased. Despite ReFAB's complexity, we found little evidence of overfitting in our calibration (*20*). For example, ReFAB predicted the AGWB of 78 validation sites randomly held out from the model-fitting process nearly as well as the calibration-training data (purple triangles and dashed line in Fig. 1B). We therefore applied ReFAB to predict AGWB

[1]Department of Biological Sciences, University of Notre Dame, South Bend, IN, USA. [2]Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, MD, USA. [3]Department of Statistics, University of California, Berkeley, Berkeley, CA, USA. [4]Department of General Education, Mount Royal University, Calgary, Alberta, Canada. [5]Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada. [6]US Geological Survey, Southwest and South Central Climate Adaptation Centers, Tucson, AZ, USA. [7]Department of Geosciences, University of Arizona, Tucson, AZ, USA. [8]Department of Forest and Wildlife Ecology, University of Wisconsin–Madison, Madison, WI, USA. [9]Department of Geography, University of Wisconsin–Madison, Madison, WI, USA. [10]Center for Climatic Research, University of Wisconsin–Madison, Madison, WI, USA.
*Corresponding author. Email: ann.raiho@gmail.com (A.M.R.); jmclach@nd.edu (J.S.M.)

over the past 10,000 years at 77 fossil pollen sites. Figure 2 shows a post hoc spatial smoothing of these estimates (*20*). The uncertainty bounds in Fig. 2B incorporate the uncertainty in the pollen-to-AGWB calibration (Fig. 1B) and also uncertainties in sediment dating and in smoothing over space and time (*20*). Despite these inevitable obstacles to quantifying biomass change from fossil assemblages, we were able to detect clear long-term trends in AGWB.

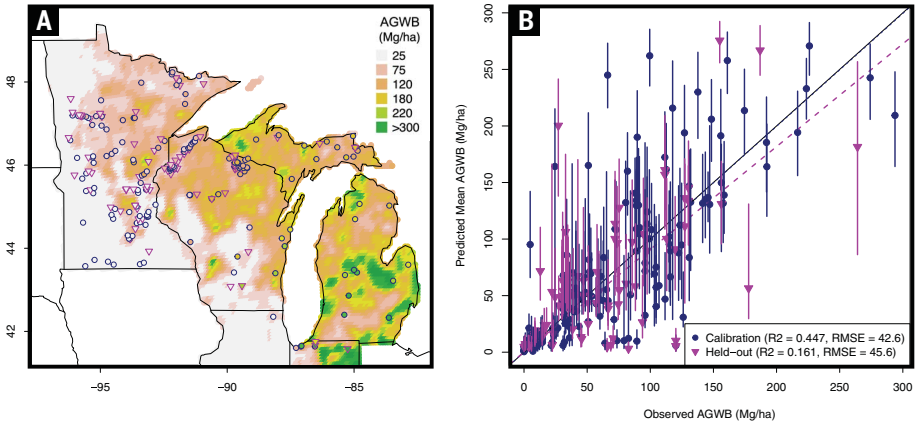


Fig. 1. AGWB at ~1840 CE. (**A**) Distribution of AGWB in the upper Midwestern region of the United States before industrial-era land clearance with location of calibration (blue circles) and held-out validation (purple triangles) pollen sites. (**B**) Model validation of ReFAB during this calibration period. Whiskers indicate 95% posterior credible intervals. Solid line indicates 1:1 line. Dashed line indicates a linear model fit forced through zero for the held-out dataset.

After an initial loss of ~450 Tg AGWB from 10,000 to 8500 years ago, AGWB increased by >1800 Tg during the subsequent 8000 years (Fig. 2B). The overall regional trend from 8000 years ago to 1850 CE was one of steady increase, with no sign of saturation in regional biomass pools. For example, we found a posterior probability of >0.975 that regional AGWB immediately preceding settlement was higher than at any time point between 10,000 and 2000 years ago [Fig. 2B (*20*)]. This result differs substantially from prior reconstructions of forest biomass in eastern Canada, which indicated little or no change in biomass over the past 6000 years (*19*). This discrepancy might reflect real differences in forest carbon dynamics between regions, but it is also possible that changes in biomass might have been missed by the simpler statistical approach taken in (*19*). The doubling of terrestrial biomass over the Holocene reported here contrasts with biome-scale reconstructions of the region, which show only modest shifts over the past 8000 years (*9*). The storage of 1800 Tg of woody biomass over 8000 years was undone in the ~150 years since the advent of modern industrial land use (Fig. 2B) (*20*), which halved biomass regionally (*24*), consistent with other

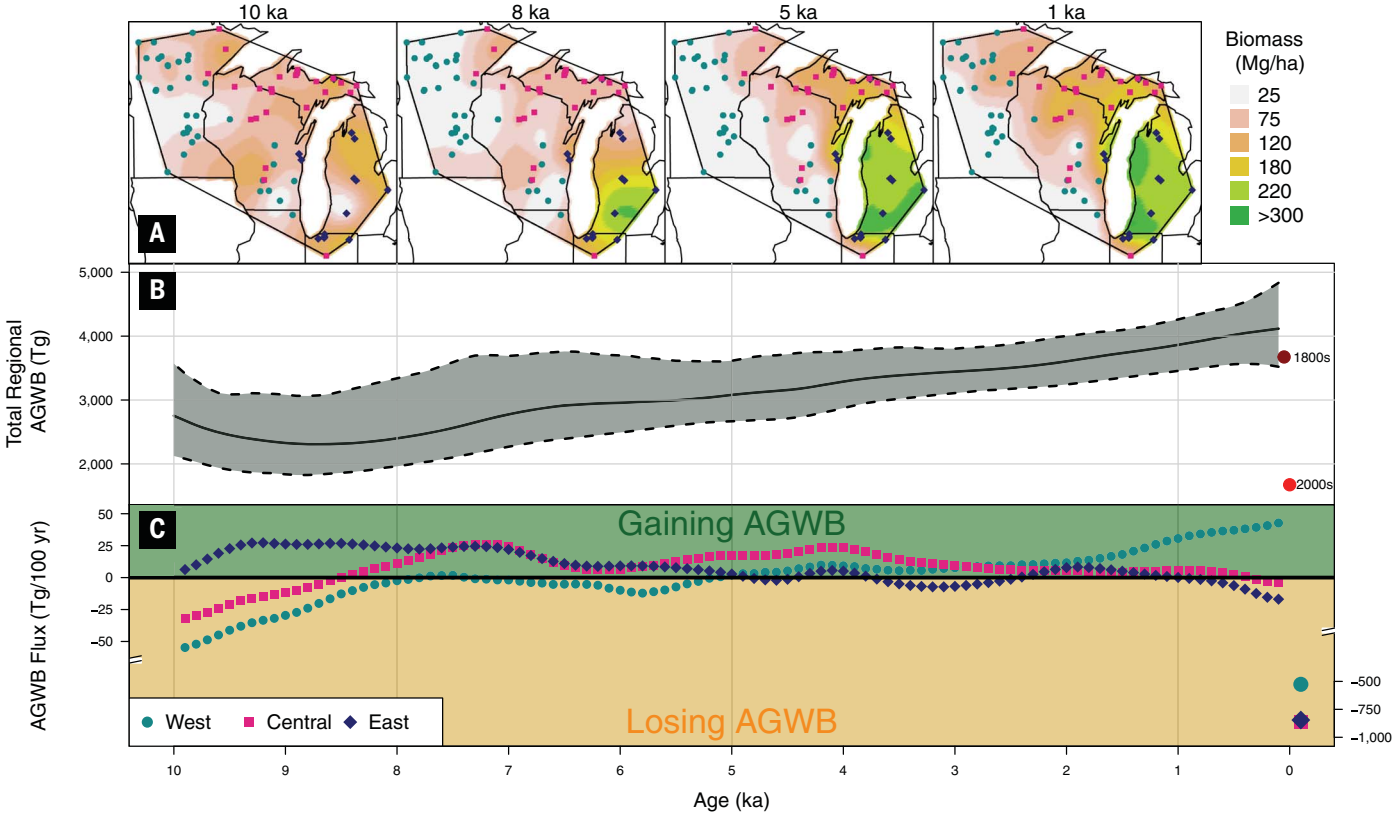


Fig. 2. Holocene changes in AGWB pools. (**A**) Spatially smoothed AGWB densities at representative Holocene time points. (**B**) Total AGWB pools across the study area with 95% uncertainty estimate: in the PIH (gray); immediately preceding forest clearance (maroon) (*21*); and currently (red) (*35*). (**C**) Century-to-century changes in total AGWB pools in regional clusters (see Fig. 4). The final time step is the difference between AGWB immediately preceding forest clearance (*21*) and currently (*35*).

estimates of the global impact of forest clearance and conversion (*11*).

We place the spatial and temporal trends in these biomass reconstructions (Fig. 2A) in the context of established understanding of Holocene changes in Midwestern vegetation (e.g., Fig. 3) (*25*–*30*) and in regional paleoclimate, broadly characterized as warming in the early Holocene (10,000 to 8000 years ago), followed by a trend toward cooler and moister climates (*26*, *31*). The fine-scale ecological processes of disturbance and demography are invisible at the resolution of this study, but they link climate and vegetation in complex ways, including lags (*27*) and asynchronous vegetation response to climate forcing (*30*). Such processes aggregate to broader trends in vegetation that can be interpreted in light of a changing macroclimate. For example, the western boundary of forest biomes retreated eastward from 10,000 to 8000 years ago and returned westward over the past 6000 years, consistent with early Holocene warming and subsequent increasing moisture [Fig. 3 (*25*)]. Within forested biomes, the boundary between coniferous and mixed coniferous/deciduous forests shifted northward in the warming climates of the early Holocene. These mixed forests generally had higher stem densities and were dominated by longer-lived and higher-biomass species. Later in the Holocene, the ranges of individual tree species expanded individualistically and asynchronously within the mixed-forest biome, reaching near-modern distributions by ~2000 years ago (*9*, *25*, *28*). The late Holocene range expansions of eastern hemlock (*Tsuga canadensis*) and American beech (*Fagus grandifolia*) [Fig. 3 (*27, 28*)] may have strongly influenced AGWB pools, because these large, long-lived species had close associations with high-biomass forests during our 19th-century calibration period (fig. S4).

A time-series clustering algorithm (*20*) identified three geographically distinct Holocene trajectories of AGWB (Fig. 4, A and B), each of which was associated with different components of the vegetation history described above. In the eastern domain (cluster E; Figs. 2 and 4), high-biomass species such as American beech expanded into a previously boreal biome in the early Holocene, sequestering the region's highest concentration of PIH woody carbon (Fig. 4, C and D). Although the north-central region (cluster C) experienced a similar biome-level transition from coniferous to mixed forests in the early Holocene, the rise in biomass at most cluster C sites occurred after this biome shift and was primarily associated with a westward range expansion of eastern hemlock within the mixed-forest biome (Fig. 4, C and D). Although hemlock declined across its range in the mid-Holocene (*29*), woody biomass at sites in cluster C was maintained by other high-biomass species (fig. S18) (*20*, *28*). Cluster W primarily includes sites in tallgrass prairie or lower-biomass forests, which consequently have low Holocene AGWB densities throughout the PIH (Fig. 4C). Biomass trends in cluster W correspond to Holocene shifts in the prairie-forest border. The eastward retreat of forests in cluster W associated with early Holocene warming and drying was the primary driver of a net loss of woody biomass across the entire study area (10,000 to 8000 years ago; Fig. 2C), and late Holocene forest expansion with the return of moisture to this region contributed to biomass accumulation, especially during the past 2000 years (Figs. 2C and 4C) (*25*, *30*).

Despite common regional climate drivers (*31*) and a smooth, "J-shaped" AGWB trajectory across the entire study area (Fig. 2B), periods of carbon storage or release in woody biomass pools are generally uncorrelated across subregions (Fig. 2C). This reflects the fact that the regional trend is an epiphenomenon driven by spatially and temporally distinct dynamics at different ecological scales. The regional loss of AGWB before 8000 years ago was primarily driven by the eastward retraction of forests in cluster W and, to a lesser extent, in cluster C (Figs. 2C and 4C), consistent with early Holocene warming and drying. These losses outpaced the coeval doubling of AGWB in cluster E, where a coniferous biome was replaced by mixed forests. Moister climates in the late Holocene (*31*) had little effect on biomass storage in cluster E (Fig. 2C), where high-biomass forests had already been established (Fig. 4C). In cluster C, closer to the dry western boundary of forests, species such as beech and hemlock expanded within the mixed-forest biome, consistent with a response to increased moisture sequestering additional woody biomass until cluster C AGWB asymptoted to near modern values by ~3000 years ago (Figs. 2C and 4C). In cluster W, increases in AGWB associated with the late-Holocene expansion of forested biomes did not occur until after 2000 years ago (Figs. 2C and 4C). Of the two primary ecological mechanisms responsible for the millennial-scale changes in woody biomass that we document here, the retraction and expansion of biome boundaries is already recognized as being an important driver of long-term terrestrial carbon budgets (*10*, *25*). The second mechanism, the range shifts of individual high-biomass tree species, has been suspected to contribute to long-term

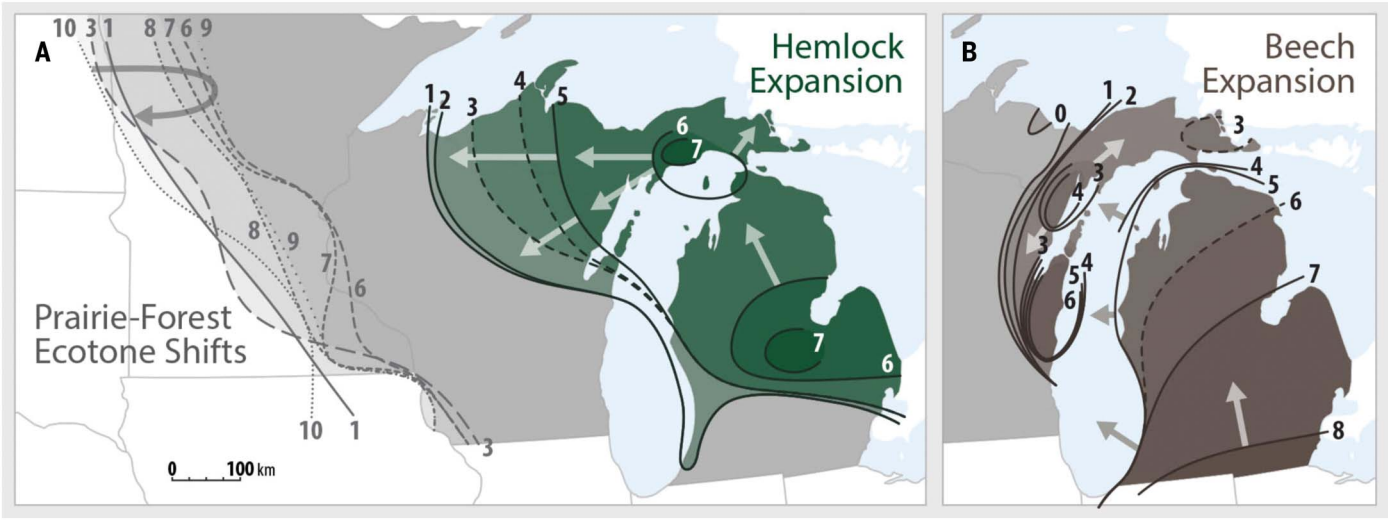


Fig. 3. Holocene vegetation shifts. (**A**) Retreat and advance of the grassland-forest ecotone (*25*) and the northwestward expansion of eastern hemlock (*27*). (**B**) Northwestward expansion of American beech (*27*).

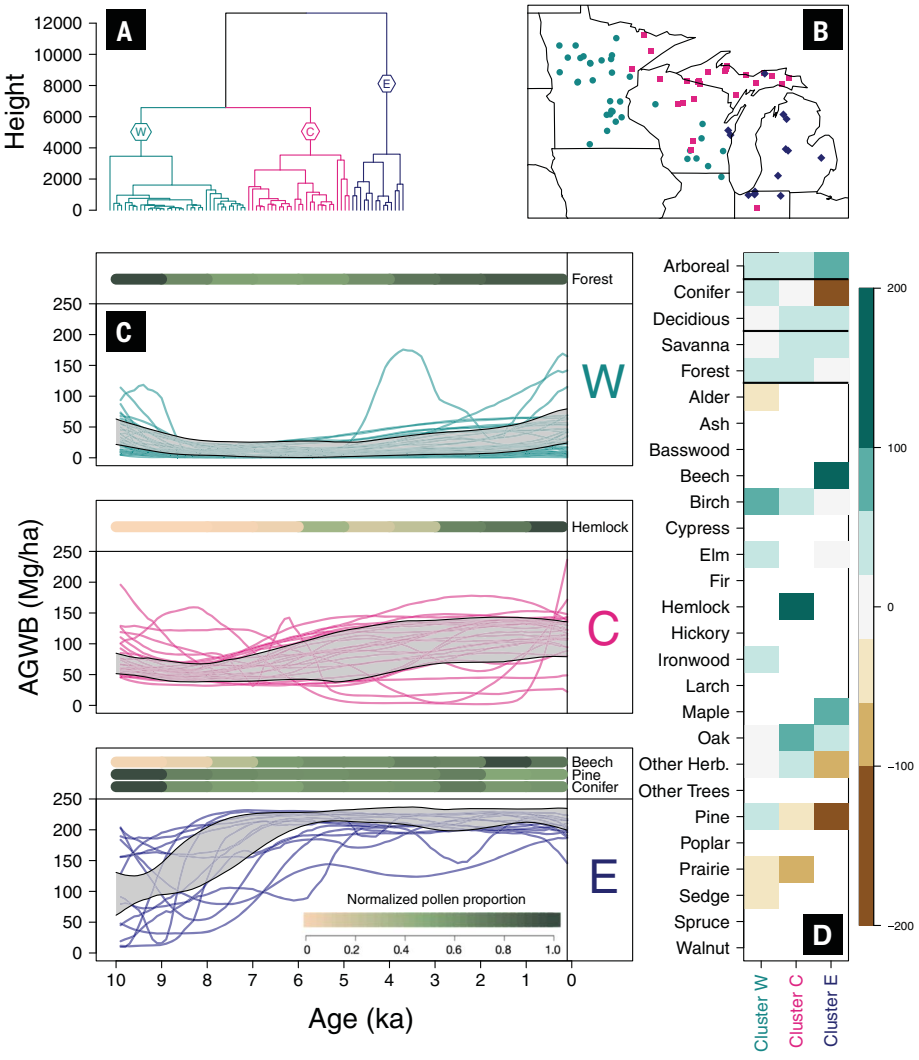


Fig. 4. Holocene trajectories of AGWB density. (**A**) Classification of 77 Holocene AGWB trajectories into three clusters: eastern (E, purple), central (C, pink), and western (W, green), where height is proportional to intergroup dissimilarity. (**B**) Site locations in the three trajectory clusters. (**C**) For each cluster, mean AGWB trajectories of each site (lines) with 50% confidence bounds across sites. (**D**) Sensitivity coefficients of AGWB to pollen abundance for individual taxa or for taxa grouped by functional or ecosystem categories (*20*).

AGWB dynamics (*13*), but is not included in biome-level Holocene biomass reconstructions (*10*–*12*).

Our finding that forest biomass accumulated over millennia carries several implications for contemporary forest and carbon modeling and management. First, these reconstructions provide context for the rapid rate of depletion of biomass pools in the industrial era: The decline in woody biomass after land clearance in each of our clusters was >10 times the rate of AGWB change in any century over the past 10,000 years (Fig. 2C). Second, the distribution of woody biomass immediately preceding the start of the industrial era (*21, 22*) is often used as a steady-state baseline for carbon cycle modelers. Our analysis shows that the late Holocene terrestrial carbon cycle in eastern North America was not in steady state. Rather, carbon pools changed slowly and continuously over the PIH, with substantial regional pools of carbon stored in woody vegetation from 8000 years ago to the time of Euro-American settlement (Fig. 2B). Third, our reconstruction of the natural pace and pattern of carbon storage and release in vegetation is a useful contribution to ongoing debates concerning whether landscapes can be managed to maximize carbon storage while also meeting conservation goals (*32*). The loss of almost 500 Tg of biomass from Midwestern vegetation during the early Holocene (Fig. 2B) demonstrates the risk posed by warming climates to carbon in forests (*33*). However, the subsequent storage of 1800 Tg of biomass in the region was largely driven by the slow areal and population expansion of high-biomass tree species. Once high-biomass species were established in our clusters E and carbon, high-biomass forests were sustained regionally for millennia. Our reconstructions confirm arguments that high-biomass species in old forests play important roles in carbon storage (*13*, *34*). Refocusing forest management toward a greater emphasis on sustainable populations of large trees could emulate the natural processes that enhanced carbon sequestration throughout the PIH and extend the time scales at which terrestrial ecosystems will continue to buffer climate change by acting as a carbon sink.

REFERENCES AND NOTES

1. S. T. Jackson, J. T. Overpeck, *Paleobiology* **26**, 194–220 (2000).
2. B. L. Otto-Bliesner *et al.*, *Geosci. Model Dev.* **10**, 3979–4003 (2017).
3. D. Purves, S. Pacala, *Science* **320**, 1452–1453 (2008).
4. A. D. Friend *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3280–3285 (2014).
5. F. H. Bormann, G. E. Likens, *Am. Sci.* **67**, 660–669 (1979).
6. D. S. Schimel, G. P. Asner, P. Moorcroft, *Front. Ecol. Environ.* **11**, 129–137 (2013).
7. S. Sitch *et al.*, *Glob. Change Biol.* **14**, 2015–2039 (2008).
8. M. Albani, D. Medvigy, G. C. Hurtt, P. R. Moorcroft, *Glob. Change Biol.* **12**, 2370–2390 (2006).
9. J. W. Williams, B. N. Shuman, T. WebbIII, P. J. Bartlein, P. L. Leduc, *Ecol. Monogr.* **74**, 309–334 (2004).
10. I. C. Prentice, S. P. Harrison, P. J. Bartlein, *New Phytol.* **189**, 988–998 (2011).
11. K.-H. Erb *et al.*, *Nature* **553**, 73–76 (2018).
12. J. R. Allen *et al.*, *J. Biogeogr.* **47**, 2073–2090 (2020).
13. O. L. Phillips *et al.*, *Surv. Geophys.* **40**, 913–935 (2019).
14. A. Ahlström, J. Xia, A. Arneth, Y. Luo, B. Smith, *Environ. Res. Lett.* **10**, 054019 (2015).
15. D. Magri, *Rev. Palaeobot. Palynol.* **81**, 313–325 (1994).
16. H. Seppä *et al.*, *Holocene* **19**, 209–220 (2009).
17. O. Blarquez, C. Carcaillet, T. M. Elzein, P. Roiron, *Holocene* **22**, 579–587 (2012).
18. C. A. Knight *et al.*, *Holocene* **31**, 814–829 (2021).
19. O. Blarquez, J. C. Aleman, *Can. J. For. Res.* **46**, 485–498 (2016).
20. See the supplemental materials.
21. C. J. Paciorek *et al.*, *PLOS ONE* **16**, e0246473 (2021).
22. J. M. Rhemtulla, D. J. Mladenoff, M. K. Clayton, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6082–6087 (2009).
23. A. Dawson *et al.*, *Quat. Sci. Rev.* **137**, 156–175 (2016).
24. J. M. Rhemtulla, D. J. Mladenoff, M. K. Clayton, *Ecol. Appl.* **19**, 1061–1078 (2009).
25. J. W. Williams, B. Shuman, P. J. Bartlein, *Global Planet. Change* **66**, 195–207 (2009).
26. J. W. Williams, B. Shuman, P. J. Bartlein, N. S. Diffenbaugh, T. WebbIII, *Geology* **38**, 135–138 (2010).
27. M. Davis, K. Woods, S. Webb, R. Futyma, *Vegetatio* **67**, 93–103 (1986).
28. S. T. Jackson *et al.*, *Quat. Sci. Rev.* **98**, 60–73 (2014).
29. R. K. Booth, S. Brewer, M. Blaauw, T. A. Minckley, S. T. Jackson, *Ecology* **93**, 1841–1852 (2012).
30. C. E. Umbanhowar Jr., P. Camill, C. E. Geiss, R. Teed, *Quat. Res.* **66**, 53–66 (2006).
31. B. N. Shuman, J. Marsicek, *Quat. Sci. Rev.* **141**, 38–51 (2016).
32. N. Seddon *et al.*, *Glob. Chang. Biol.* **27**, 1518–1546 (2021).
33. W. R. L. Anderegg *et al.*, *Science* **368**, eaaz7005 (2020).
34. S. Luyssaert *et al.*, *Nature* **455**, 213–215 (2008).
35. M. Santoro, GlobBiomass – global datasets of forest biomass (PANGAEA, 2018); https://doi.org/10.1594/PANGAEA.894711.
36. Data for: A. M1. Raiho *et al.*, 8000-year doubling of Midwestern forest biomass driven by population- and biome-scale processes, PackageId: msb-paleon.45.0, EDI Data Portal (2020); https://doi.org/10.6073/pasta/9e22f05a75e697e20c227f42c2c72d09.

37. Code for: A. M. Raiho *et al*., 8000-year doubling of Midwestern forest biomass driven by population- and biome-scale processes, Zenodo (2022); https://doi.org/10.5281/zenodo.6473397.

ACKNOWLEDGMENTS

We thank the Paleo Ecological Observatory Network (PalEON) for supporting project development; the data stewards and contributors of the Neotoma Paleoecology Database and the Public Land Survey data; J. Peters for data digitizing and archiving; and J. S. Clark and four anonymous reviewers for greatly improving the manuscript. **Funding:** This material is based upon work supported by the National Science Foundation MacroSystems Biology under grant nos. DEB-1241874, 1241868, and 1241851. **Author contributions:** A.M.R., C.J.P., and J.S.M. designed the study with conceptualizations that included A.G.D., J.W.W., and S.T.J. D.J.M. provided the original settlement vegetation dataset. A.M.R. performed the analyses with supervision from C.J.P. and J.S.M. A.M.R. and J.S.M. led the writing of the main text with revisions from all authors. A.M.R. and C.J.P. led the writing of the supplemental material text with revisions from all authors. All authors contributed to the article and approved the submitted version. **Competing interests:** The authors declare no competing interests. Any use of trade, firm, or product name is for descriptive purposes and does not imply endorsement by the US Government. **Data and materials availability:** All data are available in the supplementary materials and on the Environmental Data Initiative's data portal (*36*). All code is available on Zenodo (*37*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk3126
Materials and Methods
Figs. S1 to S18
Table S1
References (*38*–*72*)
Data S1

Submitted 15 July 2021; accepted 3 May 2022
10.1126/science.abk3126

TECHNICAL COMMENT

Comment on "The early origin of a birdlike inner ear and the evolution of dinosaurian movement and vocalization"

Romain David[1]*, Mario Bronzati[2], Roger B. J. Benson[3]

Hanson *et al.* (Research Articles, 7 May 2021, p. 601) claim that the shape of the vestibular apparatus reflects the evolution of reptilian locomotion. Using biomechanics, we demonstrate that semicircular canal shape is a dubious predictor of semicircular duct function. Additionally, we show that the inference methods used by Hanson *et al.* largely overestimate relationships between semicircular canal shape and locomotion.

Hanson *et al.* (*1*) analyzed shape variation in the inner ear of reptiles, including birds and extinct archosaurs. For the vestibular apparatus, including the semicircular canals, they found three distinct shape clusters and interpreted them as being functionally related to major locomotory categories, including quadrupedalism, bipedalism/simple flight, and agile flight. However, their interpretations are not supported by either their analyses or mathematical models of semicircular duct function. In particular, we argue that semicircular canal shape has no predictive power for inferring locomotion in reptiles, thereby refuting that its evolution can primarily be explained by locomotory shifts.

The three semicircular ducts of the inner ear are connected, endolymph-filled, toroidal membranous organs, each containing a diaphragm-like structure called the cupula (*2*). Semicircular ducts monitor head rotations and can be modeled as heavily damped torsion pendulums (*3*). The resulting biomechanical models (*2*–*4*) are central to evaluating any link between form and function of these structures. Head motion proprioception being vital (*5*), semicircular duct function is expected to attune to the spectrum of head rotations experienced by an organism (*6*), suggesting a possible link with behavior and especially locomotion (*7*).

Most comparative studies indirectly access semicircular duct structure through the enclosing bony semicircular canals, and many of them (*8*–*10*), including Hanson *et al.*, rely on geometric morphometrics to quantify canal shape. However, this widespread approach is flawed. Semicircular canal shape is fundamentally unsuited to predict behavioral signal because shape analyses aggregate morphological signal in a way that is incompatible with biomechanical models. For example, highly different semicircular duct shapes can have the exact same function, illustrated here using a two-dimensional example (Fig. 1A; labyrinths a1, b1, and c1 have different shapes but identical function). Additionally, size differences between similarly shaped semicircular ducts translate into important differences in function that are not captured by shape analysis (Fig. 1A; labyrinths a1, a2, and a3 have near-identical shapes but very different functions). Therefore, results of geometric morphometric shape analyses can be incompatible with those of functional analyses (Fig. 1, B and C).

Interpretation of form-function relationships is further complicated by a well-characterized statistical issue associated with comparative datasets (*11*): Traits of closely related species tend to be more similar simply because they have a longer duration of shared ancestry. This violates the assumption of "ordinary" (non-phylogenetic) statistical methods, specifically that data points represent independent observations, causing high rates of false positives unless data are analyzed using phylogenetic comparative methods (*11*).

Hanson *et al.* do present some phylogenetic comparative analyses (*12*), which return only minimal associations between locomotion and semicircular canal shape (*1*). For example, at various phylogenetic scales, the associations of canal shape to key locomotor traits are very weak [flight, R^2 = 0.026 (Archosauria); semi-aquatic habits, R^2 = 0.028 (Reptilia); aerial

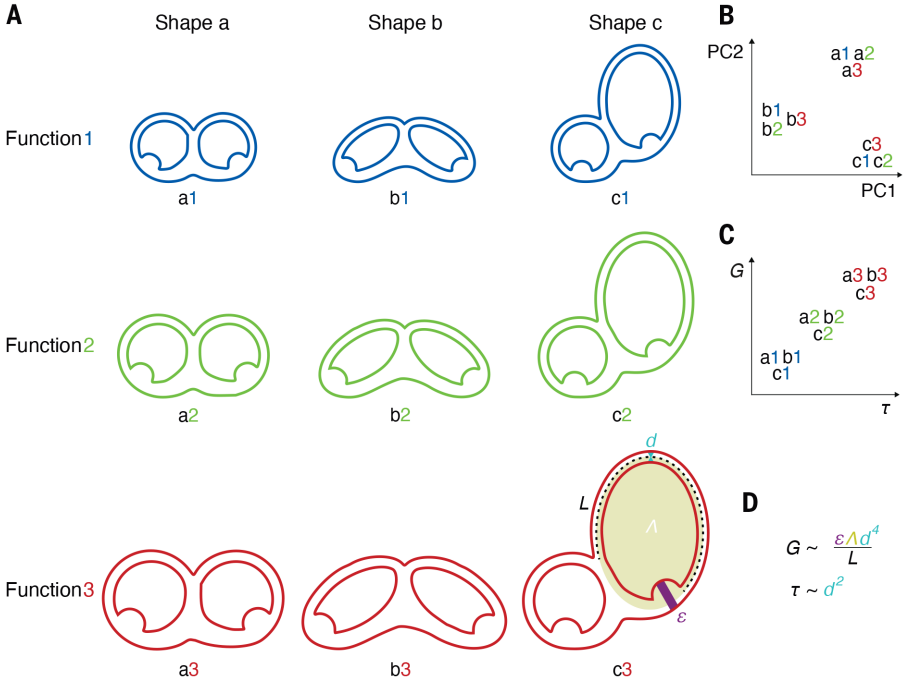


Fig. 1. Illustration of the incongruence between semicircular duct shape and function. (**A**) Diagrams representing idealized anterior and posterior semicircular duct pairs, inspired by existing mammal (a), lizard (b), and bird (c) morphologies. Duct pairs with similar colors have similar function (1, 2, or 3), measured as their response speed τ and sensitivity *G*. Duct pairs of the same column have similar shapes but different sizes. *L*, length of the slender portion of a duct; Λ, area enclosed by a duct torus; *d*, diameter of the cross section of the slender portion of a duct; ε, deflection factor of the cupula, correlated to its overall size. (**B**) Plot of a principal components analysis of the shape of duct pairs illustrated in (A). Note that specimens cluster by shape, not by function. (**C**) Plot of sensitivity versus response speed for semicircular duct pairs illustrated in (A). Note that specimens cluster by function, not by shape, and that the plots in (B) and (C) are incompatible. (**D**) Formulas used to calculate semicircular duct sensitivity and response time according to (*4*).

[1]Centre for Human Evolution Research, Natural History Museum, London, UK. [2]Department of Biology, University of São Paulo, Ribeirão Preto, Brazil. [3]Department of Earth Sciences, University of Oxford, Oxford, UK.
***Corresponding author. Email: r.david@nhm.ac.uk**

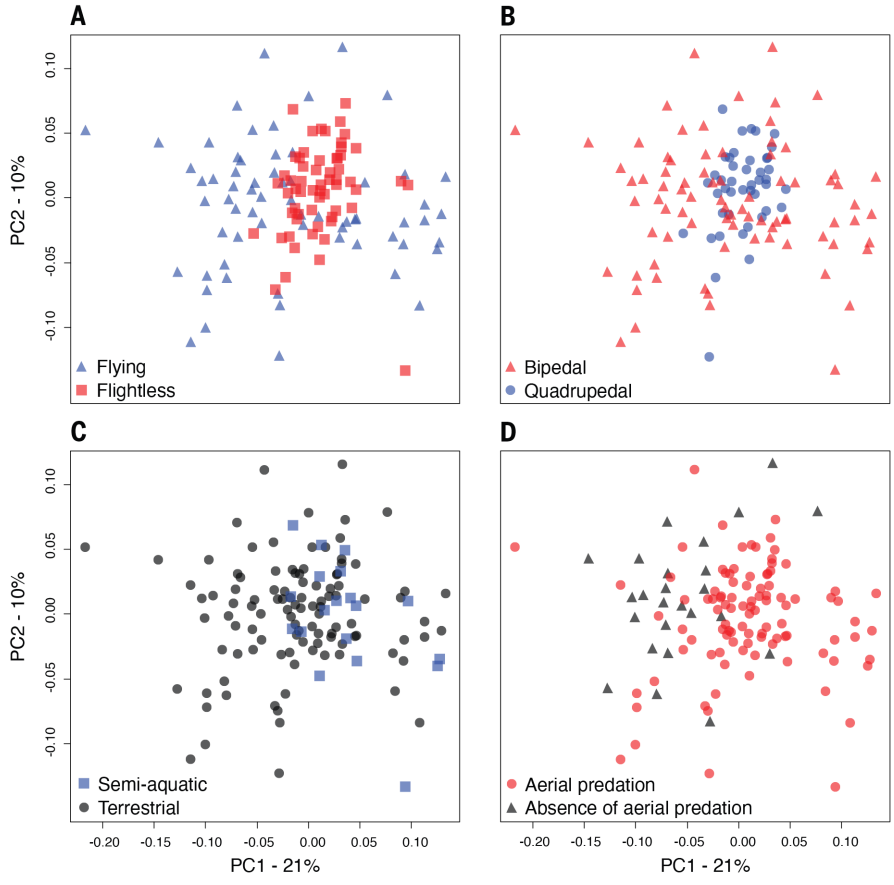


Fig. 2. Phylogenetically transformed principal components analysis of shape data and behavioral signal, showing the absence of locomotor-related clusters when shape data are corrected for phylogeny. (**A**) Flying versus flightless. (**B**) Bipedal versus quadrupedal. (**C**) Semiaquatic versus terrestrial. (**D**) Presence or absence of aerial predation. The strong overlap of behavioral categories suggests that the clusters reported by Hanson *et al*. reflect phylogenetic signal and not behavioral signal.

predation, R^2 = 0.115 (Neoaves); i.e., 2.6%, 2.8%, and 11.5% of shape variance explained, respectively; table 1 of (*1*)]. Such small portions of shape variation are expected to be functionally negligible. Indeed, according to biomechanical models (*2*–*4*), even large shape variations encompassing semicircular duct eccentricities from 0 (fully circular) to 0.87 (highly elliptical) result in maximum sensitivity differences of only 18% when centroid size is kept constant. Similar results could theoretically be achieved by increasing the overall size of any semicircular canal by ~9%, which is well within intraspecific variation (*13*).

Hanson *et al*. also present a suite of nonphylogenetic approaches that are the main justification for their conclusions but are not appropriate for comparative data. These approaches include observation of clusters in morphospace, diagrams showing semicircular canal shapes mapped onto trees with verbal interpretations, and nonphylogenetic statistical tests including linear discriminant analysis. Their linear discriminant analysis suggests that canal shapes can accurately predict the occurrence of flight (86.4 to 100% reported accuracy), semiaquatic habits (83.2 to 99.2%), and aerial predation (81.3 to 99.2%). We argue that these high values are artifacts of the failure to analytically accommodate the biasing effects of shared ancestry (*11*), which allows members of a group to share similar semicircular canal shapes, as well as similar locomotor traits, without any causal relationship (*11*). We illustrate this by simulating scores for the first 10 principal component axes of shape based on landmark data from Hanson *et al*., using Brownian motion and the tree they provided. Running linear discriminant analyses on these dummy shape scores, explicitly uncorrelated with locomotor traits, we obtained jackknifed accuracies of 91% (flight), 99% (bipedalism), 85% (semiaquatic habits),1 and 85% (aerial predation). This demonstrates that linear discriminant analysis will return a strong association of canal shape with locomotion even when none exists. To address this, we then ran linear discriminant analyses on scores obtained from a phylogenetically transformed principal components analysis (*14*) of the shape data of Hanson *et al*. We found jackknifed accuracies of 40% (flight), 44% (bipedalism), 21% (semiaquatic habits), and 20% (aerial predation). Compare these to a null model accuracy of 50% obtained when classifications are purely guesswork (Fig. 2).

Evidence that canal shape does not provide accurate predictions of locomotor traits was confirmed by another study of the semicircular canals of living and extinct archosauromorphs (*15*). That study reported multiple convergent origins of near-birdlike canal shapes in stem archosaurs (flightless, sprawling quadrupeds), early avemetatarsalians (flightless bipeds), pterosaurs (flying quadrupeds or bipeds), and the closest non-avialan dinosaur relatives of birds (e.g., *Velociraptor*; flightless bipeds). Variation in locomotor styles and agility levels among these taxa further demonstrates that labyrinth shape does not allow reliable predictions of locomotor traits in either extant or extinct species.

In summary, we argue that locomotor predictions made from semicircular canal shape are dubious at best. Given the absence of strong links between semicircular canal shape and function, it seems more likely that canal shapes are more strongly influenced by spatial and developmental constraints (*9*, *15*). For example, birds and reptiles have different braincase geometry and relative brain size, providing a nonlocomotory explanation of their labyrinth shape. Future studies of semicircular canals should instead focus on biomechanical indices of function when aiming to infer behavioral signal.

REFERENCES AND NOTES

1. M. Hanson, E. A. Hoffman, M. A. Norrell, B. S. Bhullar, *Science* **372**, 601–609 (2021).
2. R. D. Rabbitt, E. R. Damiano, J. W. Grant, in *The Vestibular System* (Springer, 2004), pp. 153–201.
3. C. M. Oman, E. N. Marcus, I. S. Curthoys, *Acta Otolaryngol.* **103**, 1–13 (1987).
4. R. David, A. Stoessel, A. Berthoz, F. Spoor, D. Bennequin, *Sci. Rep.* **6**, 32772 (2016).
5. D. E. Angelaki, K. E. Cullen, *Annu. Rev. Neurosci.* **31**, 125–150 (2008).
6. G. M. Jones, K. E. Spells, *Proc. R. Soc. London Ser. B* **157**, 403–419 (1963).
7. F. Spoor, *Courier Forschungsinst. Senckenb.* **243**, 93–104 (2003).
8. B. V. Dickson, E. Sherratt, J. B. Losos, S. E. Pierce, *R. Soc. Open Sci.* **4**, 170058 (2017).
9. R. B. J. Benson, E. Starmer-Jones, R. A. Close, S. A. Walsh, *J. Anat.* **231**, 990–1018 (2017).
10. C. Grohé, Z. J. Tseng, R. Lebrun, R. Boistel, J. J. Flynn, *J. Anat.* **228**, 366–383 (2016).
11. J. Felsenstein, *Am. Nat.* **125**, 1–15 (1985).
12. D. C. Adams, *Evolution* **68**, 2675–2688 (2014).
13. F. Spoor, F. Zonneveld, *Am. J. Phys. Anthropol.* **107** (suppl. 27), 211–251 (1998).
14. M. L. Collyer, D. C. Adams, *Methods Ecol. Evol.* **12**, 359–372 (2020).
15. M. Bronzati *et al*., *Curr. Biol.* **31**, 2520–2529.e6 (2021).

ACKNOWLEDGMENTS

Funding: Supported by the Calleva Foundation (R.D.). **Author contributions:** R.D. and R.B.J.B. conceived the study. R.D. performed analyses and biomechanical calculations. R.D., M.B., and R.B.J.B. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests.

Submitted 28 July 2021; accepted 6 June 2022
10.1126/science.abl6710

TECHNICAL RESPONSE

Response to Comment on "The early origin of a birdlike inner ear and the evolution of dinosaurian movement and vocalization"

Michael Hanson[1,2], Eva A. Hoffman[3], Mark A. Norell[3], Bhart-Anjan S. Bhullar[1,2]*

David *et al*. claim that vestibular shape does not reflect function and that we did not use phylogenetic inference methods in our primary analyses. We show that their claims are countered by comparative and direct experimental evidence from across Vertebrata and that their models are empirically unverified. We did use phylogenetic methods to test our hypotheses. Moreover, their phylogenetic correction attempts are methodologically inappropriate.

David *et al*. (*1*) claim, first, that the widely accepted evolutionary association between semicircular canal shape and locomotor behavior is unsupported by evidence; second, that our phylogenetic comparative analyses fail to show significant locomotor signal. In support of their first assertion, the authors draw upon a seldom-used modeling approach unsubstantiated by empirical data (*2*). They purport to show that ostensibly function-related metrics (referred to as "function" despite lack of biological validation) can be, under certain circumstances, similar for canals of different shapes, and that canal size has a significant effect on these metrics. Even if the above claims were to be validated in real animals, they would not falsify the hypothesis that evolutionary shape change is related to locomotor function. The close association of canal shape with locomotor function is among the best-documented anatomy/function relationships in vertebrate biology, as shown by the references listed below. Critically, there is experimental confirmation from selection studies in which the experimenters selected four replicate mouse lines for increased locomotor activity (*3*). Mass-corrected semicircular canal shape differed significantly between experimentals and controls but size did not, directly contradicting David *et al*.'s interpretations and supporting ours. The authors of the experimental work conclude that "changes in locomotor behavior over a relatively small number of generations exert considerable evolutionary influence on canal morphology" and endorse "the usefulness of 3D canal morphology in the reconstruction and potential differentiation of locomotor behaviors between closely related fossil species and particularly between time sequences showing progressive changes in locomotor behaviors over time," because "[their] results suggest that morphological differences in canal shape are most likely the result of evolutionary change" owing to vestibular morphology's "capacity to reflect nuanced differences in locomotor behavior" (*3*).

The assertion by David *et al*. that semicircular canal shape has essentially no functional relevance in locomotor control contradicts an extensive body of comparative literature, including papers they co-authored (*4*, *5*). Bronzati *et al*. (*5*), contrary to David *et al*., found, as did we, that "several locomotor traits" have "significant, independent effects on labyrinth shape: bipedal/quadrupedal gait, semi-erect/erect limb postures, and flying/non-flying locomotion." The authors preferred to ascribe the correlations to indirect causes, including developmental constraints we have called into question (*6*). A vestibular shape/locomotion relationship has been demonstrated in clades encompassing all of Vertebrata, including cyclostomes and fish (*7*); amphibians (*8*); lizards [e.g., (*9*)]; sauropterygian reptiles (*4*); archosaurian reptiles (*5*); xenarthrans [e.g., (*10*)]; rodents (*11*); and carnivorans, cetaceans, and primates [e.g., (*12*)].

David *et al*.'s statement that our phylogenetic comparative analyses returned "minimal associations" between locomotion and vestibular morphology is untrue. When summarizing our results, they ignore our central analyses: phylogenetic ANOVAs (analyses of variance) used to test for significant differences among locomotor categories [as in, e.g., (*12*)]. Contrary to David *et al*., we found that after phylogenetic correction, semicircular canal morphology was significantly different among locomotor groups such that we could distinguish quadrupeds, bipeds, simple and complex fliers, and swimmers. Our sample of extant birds allowed us to detect a significant difference between "uncomplicated" fliers and secondarily flightless taxa on the one hand and "complex" fliers on the other. We found at least six independent, qualitatively evident reversals to an ancestral bird–like vestibular geometry in association with loss or simplification of flight (Fig. 1). The reversals involve changes in canal angle and reduction in length of the anterior semicircular canal, which detects changes in pitch, the mode of rotation most important for flight (*13*).

Our remaining analyses were supplementary, conducted as in previous work (*9*, *10*). Although uncorrected analyses do risk overestimating signal owing to nonindependence stemming from common ancestry, there is no accepted way to correct for phylogenetic signal when performing analyses such as linear discriminant analysis and canonical variate analysis. David *et al*. performed a so-called phylogenetic principal components analysis (phy-PCA) using our data. Phy-PCA transforms the data to minimize "phylogenetic signal" as represented by a covariance matrix based on distances derived from a tree. Because most ecological traits, even if they have appeared convergently multiple times, are also shared among close relatives, phy-PCA runs the risk of overcorrecting and removing most "real" ecological/functional data (*14*, *15*). According to (*14*), it is often the case that "the analogy that the non-phylogenetic component is 'environmental' and the phylogenetic component is not does not hold," because functional adaptation and other processes "are, in fact, phylogenetic" and p[hy]PCA is therefore not "appropriate if adaptation of morphology is being studied, or if homoplasy is being assessed." All authors caution that these transformations, because of their inherent limitations, cannot be used to falsify hypotheses of functional evolution.

David *et al*. conclude by presenting their favored hypothesis that semicircular canal shapes are unrelated to locomotor mode and are determined instead by spatial and developmental constraints imposed by the brain and braincase. Evidence from developmental biology shows that this hypothesis is flawed. It is possible, even likely, that there is developmental integration between the brain and the ear. However, the ear placode forms very early in development and is a known source of paracrine morphogenetic signals in addition to being a significant physical component of the early head (*5*). It moreover achieves an "adult-like" form early in development (*5*), when the brain remains a simple vesicular structure and the chondrocranium has yet to appear. Therefore, we argue that all evidence suggests that the ear co-develops with, or even determines, the form of the adjacent brain and that the ear induces and directs the formation of the braincase sidewall. Far from constraining the ear, it is instead not unlikely that the brain and the skull are themselves shaped in part by selection on the semicircular canals for locomotor function.

[1]Department of Earth and Planetary Sciences, Yale University, New Haven, CT, USA. [2]Peabody Museum of Natural History, Yale University, New Haven, CT, USA. [3]Division of Paleontology, American Museum of Natural History, New York, NY, USA.
*Corresponding author. Email: bhart-anjan.bhullar@yale.edu

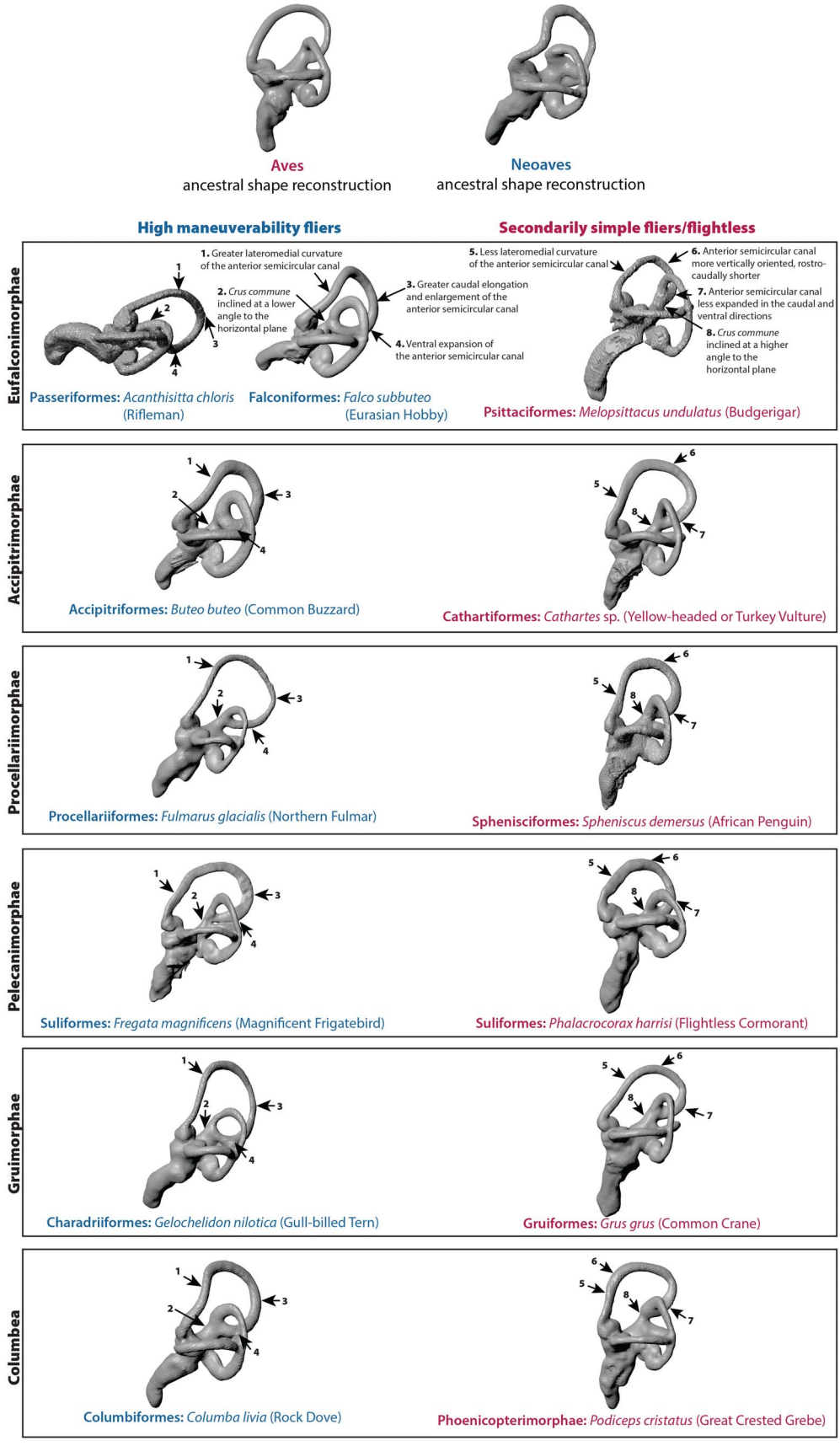

Aves
ancestral shape reconstruction
Neoaves
ancestral shape reconstruction
High maneuverability fliers
Secondarily simple fliers/flightless
Eufalconimorphae
1. Greater lateromedial curvature of the anterior semicircular canal
2. Crus commune inclined at a lower angle to the horizontal plane
3. Greater caudal elongation and enlargement of the anterior semicircular canal
4. Ventral expansion of the anterior semicircular canal
5. Less lateromedial curvature of the anterior semicircular canal
6. Anterior semicircular canal more vertically oriented, rostro-caudally shorter
7. Anterior semicircular canal less expanded in the caudal and ventral directions
8. Crus commune inclined at a higher angle to the horizontal plane
Passeriformes: Acanthisitta chloris (Rifleman)
Falconiformes: Falco subbuteo (Eurasian Hobby)
Psittaciformes: Melopsittacus undulatus (Budgerigar)
Accipitrimorphae
Accipitriformes: Buteo buteo (Common Buzzard)
Cathartiformes: Cathartes sp. (Yellow-headed or Turkey Vulture)
Procellariimorphae
Procellariiformes: Fulmarus glacialis (Northern Fulmar)
Sphenisciformes: Spheniscus demersus (African Penguin)
Pelecanimorphae
Suliformes: Fregata magnificens (Magnificent Frigatebird)
Suliformes: Phalacrocorax harrisi (Flightless Cormorant)
Gruimorphae
Charadriiformes: Gelochelidon nilotica (Gull-billed Tern)
Gruiformes: Grus grus (Common Crane)
Columbea
Columbiformes: Columba livia (Rock Dove)
Phoenicopterimorphae: Podiceps cristatus (Great Crested Grebe)

Fig. 1. Multiple reversals to "simple flier" morphology within birds.

REFERENCES AND NOTES

1. R. David, M. Bronzati, R. B. J. Benson, *Science* **376**, eabl6710 (2022).
2. R. David, A. Stoessel, A. Berthoz, F. Spoor, D. Bennequin, *Sci. Rep.* **6**, 32772 (2016).
3. H. Schutz, H. A. Jamniczky, B. Hallgrímsson, T. Garland Jr., *Evolution* **68**, 3184–3198 (2014).
4. J. M. Neenan *et al*., *Curr. Biol.* **27**, 3852–3858.e3 (2017).
5. M. Bronzati *et al*., *Curr. Biol.* **31**, 2520–2529.e6 (2021).
6. M. Hanson, E. A. Hoffman, M. A. Norell, B. S. Bhullar, *Science* **372**, 601–609 (2021).
7. S. Tamura, *Med. J. Hiroshima Univ.* **48**, 361–374 (2000).
8. H. C. Maddin, E. Sherratt, *J. Anat.* **225**, 83–93 (2014).
9. B. V. Dickson, E. Sherratt, J. B. Losos, S. E. Pierce, *R. Soc. Open Sci.* **4**, 170058 (2017).
10. G. Billet, D. Germain, I. Ruf, C. de Muizon, L. Hautier, *J. Anat.* **223**, 557–567 (2013).
11. C. Pfaff, T. Martin, I. Ruf, *Proc. R. Soc. B* **282**, 20150744 (2015).
12. C. Grohé, Z. J. Tseng, R. Lebrun, R. Boistel, J. J. Flynn, *J. Anat.* **228**, 366–383 (2016).
13. R. Dudley, *Integr. Comp. Biol.* **42**, 135–140 (2002).
14. P.D. Polly, A.M. Lawing, A.-C. Fabre, A. Goswami, *Hystrix* **24**, 33–41 (2013).
15. J. C. Uyeda, D. S. Caetano, M. W. Pennell, *Syst. Biol.* **64**, 677–689 (2015).

Submitted 3 April 2022; accepted 6 June 2022
10.1126/science.abl8181

By **Pijar Religia**

Full circle

In the exit interview after my Ph.D. defense, a professor on my thesis committee asked why I had accepted an industry job. Puzzled, I answered that it was the only offer I received; surely that was reason enough, even if the company's semiconductor business had nothing to do with my degree in biotechnology. The truth was that I had decided to pursue an industry job in my adopted country of Japan because I was tired of academia. I don't speak the language fluently, so I applied to just two companies that interviewed in English. I was relieved to land the job at the semiconductor company. I felt some anxiety about going into a totally new field, but the job would offer a new environment and keep me in Japan. I thought that was enough for me.

In the first month, my new supervisor asked me why I had chosen to work at the company, even though my degree was in a completely different field. I said I wanted to learn how industry works, and I thought this position was my only chance to do that. He didn't seem completely satisfied, and his reaction made me think more deeply about my motivations and goals. Was I just settling for whatever job came to me?

I put these concerns aside for a while. Everyone was happy with my work and assured me I was on track for a promotion. But in every quarterly meeting with my supervisor, he asked about my career plans. Every time, I felt blank. I had no vision of what I wanted to be in the company. It was like walking without a map, which felt disconcerting. Still, I carried on, hoping that maybe eventually I would come to love the work.

"I found myself reminiscing about working in academic labs."

Everything changed when my mentor at the company—a longtime employee who, in addition to my formal supervisor, served as my guide—resigned and my husband moved away for a new job. I lost my closest support system, and amid pandemic restrictions I couldn't see my husband for months. I also took on my mentor's job, which significantly increased my workload. I became lonely, depressed, and physically and mentally exhausted.

Amid these difficulties, I found myself reminiscing about working in academic labs. When I took the semiconductor job, I was burned out on academia. In the midst of the stress of completing my Ph.D., and after having spent the past decade at universities, I just wanted to get away. But in my current role, I felt I was primarily executing others' plans, leaving little room for creativity. I longed for the time when I discussed and explored ideas with my advisers, read papers, and designed my own experiments. Perhaps I had been too hasty in leaving.

Feeling lost, I talked to my family and a therapist about my growing desire to leave my job, live in the same place as my husband, and return to academia. Ultimately, I knew I needed to choose what was best for my physical and mental health. After 9 months on the job, I told my supervisor I was resigning in 3 months. He mentioned he had a hunch since our first meeting that I wouldn't stay long. He told me to get a job I really enjoy, which seemed to validate my decision.

I moved to the city where my husband was—which is also where I did my Ph.D.—and applied for postdoc positions, with the ultimate goal of securing a tenure-track job. Of course, the feeling of "what if it doesn't work again?" haunted me at first. But I reminded myself of how differently I had approached the two decisions. When I took the industry job, I had no vision of what I wanted to be. In deciding to return to academia for a postdoc, in contrast, I had reflected a lot and discussed with those dear to me. Now, I have a clearer map. I know doing research will add meaning to my life.

The director of the center where I now work is the same professor who initially asked me why I took the semiconductor job. When he interviewed me for my current position, he asked why I left the job—and I had a better answer. I told him I realized I still wanted to do academic research, and that I wanted to stay near my husband. He smiled and said, "It's good to see you again." ■

Pijar Religia is a postdoctoral researcher at Osaka University.
Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER